Complement Raising and Cluster Formation in Dutch

A Treebank-supported Investigation

Published by

LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6111

e-mail: lot@uu.nl
http://www.lotschool.nl

Cover illustration: Leen Sevens

ISBN 978-94-6093-195-6
NUR 616

KU Leuven
Faculteit Letteren
Onderzoeksgroep Computationele en
Formele Taalkunde
Centrum voor Computerlinguïstiek

Complement Raising and Cluster Formation in Dutch

A Treebank-supported Investigation

Liesbeth Augustinus

Proefschrift ingediend tot het
verkrijgen van de graad van
Doctor in de Taalkunde

Leuven, 2015

Promotiecommissie

Promotor:	Prof. dr. Frank Van Eynde
Co-promotor:	Prof. dr. Hans Smessaert
Leden:	Dr. Gosse Bouma
	Prof. dr. Jan Odijk
	Prof. dr. Jeroen van Craenenbroeck

aan mijn familie

Acknowledgements

There are many people whom I would like to thank for their support during the preparation of this thesis.

First and foremost, I would like to thank my supervisor prof. dr. Frank Van Eynde, for giving me the opportunity to write this dissertation, for the hours of inspiring discussions, and for the valuable feedback on draft versions. Without his support, this thesis would never have seen the light of day. I also want to express my gratitude to my co-supervisor prof. dr. Hans Smessaert, for the fruitful discussions and extensive comments on my writings.

Another big thank you goes out to my colleagues at the Centre for Computational Linguistics (CCL). Vincent Vandeghinste, for his co-operation on GrETEL and for helping me out with technical issues every time I got lost in the digital forest. Ineke Schuurman and Peter Dirix, for their co-operation on GrETEL, the fruitful discussions on the research presented here, and for weeding out typos in the pre-final version of this thesis. My office mates Tom Vanallemeersch and Leen Sevens, for distracting me with a sufficient amount of discussions on linguistically (ir)relevant topics when I was writing this thesis, and for designing the cover illustration.

I would like to express my gratitude to dr. Gosse Bouma and prof. dr. Jeroen van Craenenbroeck for the discussions on parts of my linguistic research, and I would like to thank prof. dr. Jan Odijk for the many suggestions and comments with respect to the development of GrETEL and tools for treebank mining.

I would like to thank the audiences of workshops and conferences where I have presented parts of the research described in this thesis, especially the audience of the past CLIN, LREC, and HPSG conferences, as well as the participants of the workshop on verb clusters. I would also like to thank the colleagues of the Department of Linguistics, with whom I have discussed my research in a more informal context.

My gratitude goes out to the the Fonds voor Wetenschappelijk Onderzoek (FWO), who funded the research presented in this thesis (G.0559.11.N.10, 2011–2015).

Many thanks to my friends and family, especially my parents Lieve and Eddy and my brother Klaas. Even though the subject of my research is somewhat mysterious to them, I could not have done this without their encouragement and endless support. And finally, thank you Jonas, for always being there.

Contents

Preface

Dutch is well-known for its verb clusters, as exemplified in the following example:

Ik denk dat ik Cecilia het nijlpaard heb zien voeren.
I think that I Cecilia the hippo have seen feed

'I think I saw Cecilia feed the hippo.'

The verbs form a cluster at the end of the clause, where they are separated from their non-verbal dependants. In addition, *Infinitivus Pro Participio* (IPP) shows up: The verb *zien* 'see' is selected by an auxiliary of the perfect, but appears as an infinitive and not as a past participle. Verb clustering and related phenomena such as the IPP effect have fascinated researchers for decades, as indicated by the abundant literature that is available within descriptive, theoretical and corpus linguistics. Still, many questions with respect to the description and the analysis of Dutch verb clusters remain unanswered. The research presented in this thesis addresses a number of these questions, and investigates how authentic language examples obtained from corpora can be an added value for a theoretical analysis of Dutch verb clusters.

This dissertation is organized into three parts: a literature study (part I), a corpus study (part II), and a theoretical analysis (part III).

Part I considers how verb clusters are described and analysed in the **descriptive and theoretical literature**. Chapter 1 gives a definition of verb clusters based on the literature, and points out the phenomena that are typically related to cluster formation. The most important ones include the IPP effect, the interruption of the cluster by non-verbal elements, and word order variation within the cluster. Chapter 2 sketches the analysis of verb clusters in transformational grammar, as the first theoretical analyses of verb clusters were described in that framework. Moreover, the terminology used in descriptive and theoretical accounts in other frameworks is often based on the transformational work on verb clusters. Chapter 3 provides an overview of the most influential monostratal analyses of verb clusters. The focus is on the treatments

formulated within Head-driven Phrase Structure Grammar (HPSG), as this framework is also used to for the new analysis proposed in part III.

The literature study addresses the following questions:

- What is the set of Dutch clustering verbs?
- In which cases is clustering obligatory and in which cases is it optional?
- What is the link between cluster formation and the IPP effect?
- What types of word order variation can be observed in Dutch verb clusters?
- What are the conditions on *cluster creeping*, i.e. the interruption of the cluster by non-verbal elements?

Part II presents **a corpus-based investigation** of verb clusters. By consulting *treebanks*, i.e. text corpora enriched with syntactic annotations, it will be investigated whether and how often the phenomena described in part I occur in non-elicited language data. Chapter 4 presents the data and the methodology used for the corpus study. Chapter 5 describes and discusses the results of the treebank investigation. The main topics that will be addressed are the word order variation observed in the data, the identification of the clustering verbs, the occurrence of IPP and cluster creeping. Special attention goes out to constructions with a *te*-infinitive, as they are often neglected in studies on verb clusters.

Part III presents a **new analysis** of Dutch verb clusters, formulated in HPSG. In chapter 6 it will be demonstrated that the current HPSG analyses do not adequately analyse Dutch verb clusters. An alternative analysis will be proposed that deals with those issues. It heavily relies on the empirical observations obtained from the treebanks. Chapter 7 illustrates how the analysis proposed in chapter 6 extends to the analysis of other phenomena, especially adposition stranding.

Part I

Literature Study

CHAPTER 1

Descriptive study

This chapter provides an overview of how the formation of verb clusters is treated in descriptive grammars. Before turning to cluster formation, the Dutch sentence structure will be discussed in section 1.1. Section 1.2 presents a general definition of Dutch verb clusters and outlines the phenomena related to verb clustering that will be discussed in further detail in the following sections, i.e. the set of clustering verbs (section 1.3), the IPP effect (section 1.4), the third construction (section 1.5), cluster creepers (section 1.6), and word order variation (section 1.7). Next, the form of the verbal complement of clustering verbs will be considered (section 1.8), as well as the position of the non-verbal arguments (section 1.9). Section 1.10 focusses on the difference between verbs that are part of the verb cluster and verbs that are in the Mittelfeld or the Nachfeld. Section 1.11 sketches verb cluster formation in German. Section 1.12 concludes. The examples provided in this chapter are based on the literature, or constructed analogous to the examples provided by the authors under discussion.

1.1 Dutch sentence structure

There are two fixed positions in the Dutch sentence. Those positions are also known as *poles* or *sentence brackets*, which serve as orientation points for the other elements in the sentence. In example (1.1a) the finite verb *heeft* 'has' occupies the first pole, while the past participle *gedronken* 'drunk' is in the second pole. In example (1.1b) the complementizer *dat* 'that' takes up the first pole, while the verbal elements *beschouwd* 'considered' and *wordt* 'be' occupy the second pole. (1.1b) shows that the second

pole may consist of multiple elements, but it can also be empty, as in example (1.1c) (Haeseryn et al. 1997: 1225–1226).

(1.1) a. Zijn zus *heeft* altijd al graag een glas wijn *gedronken*.
his sister has always already gladly a glass wine drunk
'His sister has always enjoyed a glass of wine.'

b. Het blijkt *dat* hij zowat overal ter wereld als een groot schrijver
it seems that he almost everywhere in the world as a great writer
beschouwd wordt.
considered is
'It seems that he is considered to be a great writer almost all over the world.'

c. Zijn zus *drinkt* graag wijn.
his sister drinks gladly wine
'His sister likes drinking wine.'

The poles divide sentences into *topological fields*: the *Vorfeld* (lit. 'pre-field') is the part before the first pole, the *Mittelfeld* (lit. 'middle field') is the part between the poles, and the *Nachfeld* (lit. 'after-field') refers to what follows after the second pole.

In verb-second sentences with only one verb, such as (1.1c), the second pole is empty. In verb-first sentences the finite verb occupies the first pole (1.2).

(1.2) a. *Heeft* zijn zus altijd al graag een glas wijn *gedronken*?
has his sister always already gladly a glass wine drunk
'Has his sister always enjoyed a glass of wine?'

b. *Drinkt* zijn zus graag wijn?
drinks his sister gladly wine
'Does his sister like drinking wine?'

Verb-first and verb-second constructions are canonically referred to as verb-initial constructions, while sentences with a complementizer in the first pole and one or more verbs in the second pole are referred to as verb-final constructions.

In constructions with multiple verbs at the second pole, the verbal sequence is called the *werkwoordelijke eindgroep* (lit. 'verbal end group') or *verb cluster*. The verbs at the second pole form a *cluster*, while the non-verbal elements appear before or after the second pole (1.3a–b). Nominal and adjectival constituents typically occupy a position in the Mittelfeld. Subordinate clauses typically appear in the Nachfeld, whether they are finite (e.g. clauses introduced by *dat* 'that') or non-finite (e.g. clauses introduced by *om* 'for'). Prepositional constituents may appear in either the Mittelfeld or the Nachfeld.

Canonically no non-verbal elements are allowed in the verb cluster (Haeseryn et al. 1997: 1355). The intrusion of non-verbal material in the verb clusters yields an ungrammatical sentence (1.3c).

(1.3) a. Zij beweerde dat zij het gisteren aan de juf *had gezegd*.
she claimed that she it yesterday to the teacher had said

b. Zij beweerde dat zij het gisteren *had gezegd* aan de juf.
she claimed that she it yesterday had said to the teacher
'She claimed that she told the the teacher about it yesterday.'

c. * Zij beweerde dat zij het gisteren *had* aan de juf *gezegd*.
she claimed that she it yesterday had to the teacher said

Although the impenetrability of the verb cluster is the norm in most constructions, there are some exceptions, see section 1.6.

1.2 What is a verb cluster?

In this thesis the notion *verb cluster* will be used for the sequence of two or more verbs occurring in the second pole. An example of a verb cluster consisting of two verbs was given in (1.3). The examples in (1.4) present some constructions containing longer clusters.

(1.4) a. ... dat hij haar gisteren eindelijk heeft$_1$ kunnen$_2$ ontmoeten$_3$.
... that he her yesterday finally has can meet
'... that he was finally able to meet her yesterday.'

b. ... dat zij hem zou$_1$ moeten$_2$ hebben$_3$ gezien$_4$.
... that she him would must have seen
'... that she should have seen him.'

Verb clusters are ordered in two ways. One concerns the **order of selection**. For instance in (1.4a), the finite verb *heeft* 'has' selects the bare infinitive *kunnen* 'can', which in turn selects the infinitive *ontmoeten* 'meet'. The hierarchical order is typically indicated by numbers, the hierarchically highest verb being 1.

The second way in which the sequences are ordered is the **linear order**. Canonically, the linear order of the verbs coincides with the hierarchical order, i.e. the verbs to the left select the verb to the right, as is the case in (1.4). Alternative orders are also possible, though, as shown in (1.5).

(1.5) a. ... dat zij hem zou$_1$ moeten$_2$ gezien$_4$ hebben$_3$.
... that she him would must seen have

b. ... dat zij hem zou$_1$ gezien$_4$ moeten$_2$ hebben$_3$.
... that she him would seen must have

c. ... dat zij hem gezien$_4$ zou$_1$ moeten$_2$ hebben$_3$.
... that she him seen would must have
'... that she should have seen him.'

In constructions with a past participle, such as (1.5), the participle may occur more to the front. In sentences like (1.6) the finite verb may occur as the first or as the last element of the cluster:

(1.6) a. Hij zei dat hij de chauffeur had$_1$ willen$_2$ laten$_3$ blijven$_4$ wachten$_5$.
he said that he the driver had want let stay wait
'He said that he had wanted to let the driver wait.'

b. Hij zei dat hij de chauffeur willen$_2$ laten$_3$ blijven$_4$ wachten$_5$ had$_1$.
he said that he the driver want let stay wait had
'He said that he wanted to let the diver wait.'

There are many variations on the linear order of Dutch verb clusters. The possible word orders depend on several factors, such as the type of construction involved (e.g. clusters with bare infinitives only versus clusters containing a past participle) or the dialect of the speaker. Word order variation will be demonstrated to be a diagnostic for clustering in section 1.7.

Besides word order variation within the cluster there is another diagnostic that is often used to differentiate clustering from non-clustering constructions, i.e. **Infinitivus Pro Participio** or the IPP effect. IPP refers to the occurrence of an infinitive where one would expect a past participle, i.e. if a verb is selected by an auxiliary of the perfect, as in (1.4a), (1.6), and (1.7).

(1.7) ... dat Tom haar heeft zien/*gezien dansen.
... that Tom her has see.IPP/see.PSP dance
'... that he has seen her dance.'

Section 1.4 describes the phenomenon in further detail and shows that verbs that can appear in IPP constructions are typical clustering verbs. In section 1.7 it will be shown how IPP constructions differ from non-clustering constructions with respect to word order, and in section 1.10 it will be argued that the occurrence of IPP is an important criterion to decide whether a verb is selected in the cluster or not.

While the term *verb cluster* will be used for the entire sequence of verbs in the second pole, the term *clustering verb* will be used for verbs which can select another verb within a verb cluster. An example is *gaan* 'go' in (1.8). The final verb in the hierarchical chain is the *main verb*. This verb is a part of the verb cluster, which means it occurs in the second pole, but it is not a clustering verb itself, as it does not select another verb in the cluster, cf. *slapen* 'sleep' in (1.8).

(1.8) Hij zegt dat de jongen te laat gaat$_1$ slapen$_2$.
he said that the boy too late goes sleep
'He says that the boy goes to bed too late.'

The examples with longer verb clusters in (1.4) and (1.5) indicate that a verb cluster can contain multiple clustering verbs, as one clustering verb can select another clustering verb.

The notion *clustering verb* actually refers to the clustering *use* of verbs, since the majority of those verbs can be used as main verbs as well. In example (1.8), *gaan* 'go' is used as a clustering verb, whereas in (1.9) it is a main verb.

(1.9) a. Die jongen gaat graag naar school.
that boy goes gladly to school
'That boy likes to go to school.'

b. Ik wil hem niet laten gaan.
I want him not let go
'I don't want to let him go.'

While the examples in this section contain clear-cut examples of verb clusters, it will be shown in the following sections that verb clusters are not always uniformly defined, nor is there a complete consensus on the set of verbs that can appear in verb clusters.

1.3 Clustering verbs

Haeseryn et al. (1997) make a distinction between *groepsvormende werkwoorden* (lit. 'group-forming verbs') or *clustering* verbs, *niet-groepsvormende werkwoorden* (lit. 'non-group-forming verbs') or *non-clustering* verbs, and *niet-verplicht groepsvormende werkwoorden* (lit. 'non-obligatory group-forming verbs' or *optional clustering verbs*). An example of a verb that obligatorily clusters if it selects another verb is *willen* 'want'

in (1.10). The verbal complement is always selected in the second pole (1.10a); if it appears in the Nachfeld, the construction is ungrammatical (1.10b).

(1.10) a. Hij zei dat hij de kraanvogels graag *wilde fotograferen*.
he said that he the cranes gladly wanted take-photographs

b. * Hij zei dat hij *wilde* graag de kraanvogels *fotograferen*.
he said that he wanted gladly the cranes take-photographs
'He said that he would like to take a photograph of the cranes.'

Other verbs do not cluster, such as *merken* 'notice' in (1.11). In contrast to *willen*, *merken* always selects its verbal complement in the Nachfeld (1.11a), as selection in the second pole yields an ungrammatical construction (1.11b) (Haeseryn et al. 1997: 952).

(1.11) a. Ze zeggen dat hij toen pas *merkte* een fout gemaakt te hebben.
they say that he then just noticed a mistake made to have
'They say that only then he noticed that he had made a mistake.'

b. * Ze zeggen dat hij toen pas een fout gemaakt *merkte* te hebben.
they say that he then just a mistake made noticed to have

A third group of verbs optionally clusters, such as *proberen* 'try' in (1.12). The clustering construction is given in (1.12a), whereas an example without verb cluster is provided in (1.12b) (Haeseryn et al. 1997: 950–952).

(1.12) a. Hij zei dat hij de kraanvogels *probeerde te fotograferen*.
he said that he the cranes tried to take-photographs

b. Hij zei dat hij *probeerde* de kraanvogels *te fotograferen*.
he said that he tried the cranes to take-photographs
'He said that he tried to take photographs of the cranes.'

The set of verbs with a clustering use is limited. The possible candidates are verbs selecting a verbal complement. Haeseryn et al. (1997: 1077–1081) provide an alphabetical list of such verbs. For each verb, it is indicated whether the verb is clustering or non-clustering, and which type of verbal complement it selects (bare infinitive, *te* infinitive, past participle or *aan het* infinitive). If the verb selects an infinitive, it is also specified whether the implied subject of the infinitive is identified with the subject or the object of the selecting verb. For instance, the implied subject of *leren* is identified with the subject of the embedded infinitive if it has the meaning 'learn' (1.13a), but it is identified with the object of the embedded infinitive if it has the meaning 'teach' (1.13b). In both cases, *leren* can be used in clustering and non-clustering constructions according to Haeseryn et al. (1997).

(1.13) a. Ik heb leren roeien.
I have learn row
'I have learned to row.'

b. Hij heeft me leren roeien.
He has me teach row
'He has taught me to row.'

While the set of verbs provided in Haeseryn et al. (1997) is a useful starting point, the definition of clustering verbs and/or clustering constructions is broader in Haeseryn et al. (1997) compared to other treatments of Dutch verb clusters. In section 1.10, some constructions will be discussed that are better not analysed as instances of verb clustering (e.g. constructions with predicative infinitives).

1.4 Infinitivus Pro Participio

If a main verb occurs in combination with an auxiliary of the perfect, it appears as a past participle, for example *gehoord* 'heard' in (1.14a). By contrast, some verbs appear as an infinitive if they are put in the perfect tense and if they select another infinitive, cf. *horen* 'hear' in (1.14b) (Haeseryn et al. 1997: 954). This phenomenon is known as *Infinitivus Pro Participio* (IPP) or *Ersatzinfinitiv* (lit. 'substitute infinitive').

(1.14) a. Jan heeft me wel *gehoord*.
Jan has me rather hear.PSP
'Jan did hear me.'

b. Jan heeft me *horen* zingen.
Jan has me hear.IPP sing
'Jan has heard me sing.'

c. * Jan heeft me *gehoord* zingen.
Jan has me hear.PSP sing

It is possible to differentiate between constructions that obligatorily show the IPP effect (1.14), constructions that optionally show IPP (1.15), and constructions in which IPP is not possible (1.16).

(1.15) a. De politie heeft de snelheidsmaniak *proberen* in te halen.
the police has the speed-maniac try in to overtake

b. De politie heeft *geprobeerd* de snelheidsmaniak in te halen.
the police has tried the speed-maniac in to overtake
'The police has tried to overtake the speed merchant.'

(1.16) a. * Hij heeft hem het raampje dicht *vragen* te doen.
he has him the window.DIM closed ask to do

b. Hij heeft hem *gevraagd* het raampje dicht te doen.
he has him asked to window.DIM close to do
'He asked him to close the little window.'

The examples in (1.15) and (1.16) show that the IPP effect is an indication of clustering: The sentences with IPP contain a verb cluster, whereas the examples without IPP do not. Therefore, verbs that optionally show IPP such as *proberen* 'try' in (1.15) can be used in clustering and non-clustering constructions (Haeseryn et al. 1997: 957). In Dutch the set of IPP verbs, i.e. verbs that obligatorily or optionally appear as IPP in the perfect tense, is a subset of the clustering verbs. The perfect and passive auxiliaries, for instance, are clustering verbs but do not appear as IPP.

There are several lists of IPP verbs in the literature, but none is claimed to be exhaustive. Moreover, for some verbs the authors disagree on their IPP status (i.e. whether the verbs are obligatory, optional or no IPP verbs). Table 1.1 presents a list of IPP verbs based on Rutten (1991), Klooster (2001), and IJbema (2002).[1] The verbs indicated with a '+' obligatorily occur as IPP, verbs indicated with a '–' cannot occur in IPP constructions, and verbs indicated with a '+/–' optionally occur as IPP verbs according to the author(s) mentioned in the columns. If a source does not mention the behaviour of a certain verb regarding IPP, the field is left blank. The top part of the table lists the verbs which the authors agree upon regarding IPP, while the bottom part of the table lists the verbs which were labelled differently by at least one author. The latter category also contains the verbs that were mentioned in only one source.

Lemma	**Rutten**	**Klooster**	**IJbema**
(be)horen 'ought to'	+	+	+
blijven 'stay, remain'	+	+	+
dienen 'be obliged to'	+	+	+
doen 'do, make'	+	+	+
gaan 'go, will'	+	+	+

continued on next page

[1] While Haeseryn et al. (1997) provide a list of verbs selecting a verbal complement, it is not explicitly mentioned for all those verbs whether they occur in IPP constructions or not. The verbs listed in Rutten (1991) and IJbema (2002) are taken from their list of 'Verb Raising verbs'. As will be shown in chapter 2, the Verb Raising verbs appear as IPP if they are put in the perfect tense (see also IJbema 2002: 66). Note that the auxiliaries of the perfect are also Verb Raising verbs, but they cannot appear as IPP.

Lemma	Rutten	Klooster	IJbema
(be)hoeven 'need to'	+	+	+
horen 'hear'	+	+	+
komen 'come'	+	+	+
kunnen 'can, be able to'	+	+	+
laten 'let'	+	+	+
liggen 'lie'	+	+	+
lijken 'seem'	+		+
lopen 'walk'	+		+
moeten 'must, have to'	+	+	+
mogen 'may, be allowed to'	+	+	+
staan 'stand'	+	+	+
vinden 'find'	+	+	+
voelen 'feel'	+	+	+
weten 'know (how to)'	+	+	+
willen 'want'	+	+	+
zien 'see'	+	+	+
zien 'manage'	+	+	+
zitten 'sit'	+	+	+
zullen 'will'	+	+	+
beginnen 'begin'	+/–	+/–	+/–
helpen 'help'	+/–	+/–	+/–
leren 'learn, teach'	+/–	+/–	+/–
menen 'mean, think'	+/–		+/–
pogen 'try'		+/–	+/–
proberen 'try'	+/–	+/–	+/–
trachten 'try'	+/–	+/–	+/–
wagen 'dare'	+/–	+/–	+/–
blijken 'appear'	+	–	+
denken 'think'			+/–
dreigen 'threaten'	+/–		–
durven 'dare'	+/–	+	+/–
hangen 'hang'			+
hopen 'hope'	–	+/–	–
plegen 'be used to'	+		+
ruiken 'smell'			+

continued on next page

Lemma	Rutten	Klooster	IJbema
schijnen 'seem'	+	–	+
vermogen 'be able to'			–
vrezen 'fear'	–	+/–	–
weigeren 'refuse'	+/–	–	+/–
wensen 'wish'	–	+/–	+/–
wezen 'be in the process of'		+	

Table 1.1: Dutch IPP verbs

The fact that several authors disagree on the set of verbs that can occur as IPP indicates that an empirical investigation is necessary in order to identify the Dutch IPP verbs.

The IPP phenomenon is characteristic for Dutch, but it also occurs in some other West Germanic languages, such as German and Afrikaans. Schmid (2005) compares IPP in seven languages and dialects. She distinguishes eight verb classes (e.g. modals, perception verbs, ...) that occur as IPP in at least one of the languages. The set of languages and dialects showing the IPP effect is often compared to languages with verb clusters that do not show the IPP effect. It is reported that the IPP effect does not occur in languages and dialects with a strictly descending word order in verb clusters, and that it does not occur in languages that do not mark past participles with (a variant of) the prefix *ge-*, such as Frisian. For a discussion on the distribution and potential triggers of the IPP effect, see amongst others Hoeksema (1980), Schmid (2005), and Zwart (2007).

A related question is whether IPP verbs are genuine infinitives, or rather participles in disguise. Some authors treat the IPP form as a past participle, see for instance Hinterhölzl (1999: 159–166) and Plank (2000). This assumption will not be taken up here. Since IPP verbs syntactically behave as true infinitives rather than as past participles (cf. chapter 5), they will also be analysed as such.

1.5 The Third Construction

There is a set of verbs that seems to occur in clustering constructions without appearing as IPP. Haeseryn et al. (1997) define that set of *quasi-clustering verbs* as a subset of

the optionally clustering verbs. Those verbs can occur in constructions that seem to be somewhere in between clustering and non-clustering, cf. *proberen* 'try' in (1.17a).

(1.17) a. Ze hadden al vaker duidelijk *geprobeerd* te maken dat er
they had already more-often clear tried to make that there
niets van klopte.
nothing of agreed

b. Ze hadden al vaker duidelijk *proberen* te maken dat er
they had already more-often clear try to make that there
niets van klopte.
nothing of agreed
'They had frequently tried to make it clear that nothing was correct.'

In example (1.17a) the verbs seem to cluster in a similar fashion as in (1.17b), since no non-verbal material occurs between the verb forms. The only difference between both sentences is that the IPP effect does not show up in (1.17a), but that the past participle is used.[2] A contiguous string of verbs arises because (one or more) arguments belonging to the embedded clause are realized in the matrix clause, cf. *duidelijk* 'clear' in (1.17a). The construction in (1.17a) is also known as the *third construction*, as it is a construction that seems to hold the middle between clustering and non-clustering constructions. According to Haeseryn et al. (1997: 952), the third construction occurs more often in Netherlandic Dutch than in Belgian Dutch.

Note that sentences like (1.18) are also instances of the third construction. In contrast to (1.17a) not all arguments of the embedded clause appear in the matrix clause, which means the third construction does not always show a verbal sequence (Klooster 2001: 243).

(1.18) Je hebt Jan *verzuimd* zijn geld terug te geven.
you have Jan failed his money back to give
'You failed to give John's money back.'

Rutten (1991: 78–79), Klooster (2001),[3] and IJbema (2002: 152–153) provide a list of verbs that can occur in the third construction, see Table 1.2. Verbs that may occur in the third construction but not in IPP constructions according to the author(s)

[2]This distinction will turn out to be too simplistic. In section 1.7 a more detailed account of the differences between both constructions will be presented.

[3]Klooster (2001: 246–247) lists several verbs with respect to their occurrence as IPP, but he only mentions for a subset of those verbs whether they also occur in the third construction or not (see Klooster 2001: 253–255).

mentioned in the columns are indicated as '+3, –IPP'. The verbs that can occur in the third construction as well as in IPP constructions are indicated as '+3, +IPP', whereas '–3, +IPP' is used to indicate obligatory IPP verbs, i.e. verbs that cannot occur in the third construction according to one of the authors.[4] '–3, –IPP' is used for verbs that do not occur as third construction verbs and not as IPP verbs. If one of the authors does not mention the behaviour of a certain verb regarding the third construction, the field is left blank. Similar to Table 1.1, the data are ordered according to whether there is agreement between the authors or not.

Lemma	**Rutten**	**Klooster**	**IJbema**
beginnen 'begin'	+3, +IPP		+3, +IPP
durven 'dare'	+3, +IPP		+3, +IPP
menen 'mean, think '	+3, +IPP		+3, +IPP
pogen 'try'		+3, +IPP	+3, +IPP
proberen 'try'	+3, +IPP	+3, +IPP	+3, +IPP
trachten 'try'	+3, +IPP	+3, +IPP	+3, +IPP
wagen 'dare'	+3, +IPP	+3, +IPP	+3, +IPP
aanraden 'advise'	+3, –IPP		+3, –IPP
adviseren 'advise'		+3, –IPP	+3, –IPP
begeren 'desire'	+3, –IPP		+3, –IPP
beloven 'promise'	+3, –IPP		+3, –IPP
beogen 'aim at'	+3, –IPP		+3, –IPP
besluiten 'decide'	+3, –IPP	+3, –IPP	+3, –IPP
bevelen 'order'		+3, –IPP	+3, –IPP
beweren 'claim'	+3, –IPP		+3, –IPP
dwingen 'force'	+3, –IPP	+3, –IPP	+3, –IPP
eisen 'demand '	+3, –IPP		+3, –IPP
gebieden 'command'		+3, –IPP	+3, –IPP
gelasten 'order'		+3, –IPP	+3, –IPP
geloven 'believe'	+3, –IPP		+3, –IPP
hopen 'hope'	+3, –IPP		+3, –IPP
opdragen 'order'	+3, –IPP		+3, –IPP
verbieden 'forbid'	+3, –IPP	+3, –IPP	+3, –IPP
vergeten 'forget'	+3, –IPP		+3, –IPP
verlangen 'demand'	+3, –IPP		+3, –IPP

continued on next page

[4]Those verbs are also included in Table 1.1.

Lemma	Rutten	Klooster	IJbema
verleren 'forget'	+3, –IPP		+3, –IPP
vermijden 'avoid'	+3, –IPP		+3, –IPP
verplichten 'oblige'	+3, –IPP		+3, –IPP
vertellen 'tell'	+3, –IPP		+3, –IPP
verwachten 'expect'	+3, –IPP		+3, –IPP
verzoeken 'request'		+3, –IPP	+3, –IPP
verzuimen 'neglect'	+3, –IPP	+3, –IPP	+3, –IPP
voorstellen 'propose'	+3, –IPP		+3, –IPP
vragen 'ask'		+3, –IPP	+3, –IPP
vrezen 'fear	+3, –IPP		+3, –IPP
zeggen 'say'	+3, –IPP		+3, –IPP
bezweren 'swear to'		+3, –IPP	
denken 'think, plan'	+3, –IPP		+3, +IPP
dreigen 'threaten'	+3, +IPP		+3, –IPP
smeken 'beg'		+3, –IPP	
suggereren 'suggest'		+3, –IPP	
verklaren 'declare'			+3, –IPP
vermogen 'be able to'			+3, –IPP
weigeren 'refuse'	+3, +IPP	+3, –IPP	+3, +IPP
wensen 'wish'	+3, –IPP		+3, +IPP
helpen 'help'	+3, +IPP	–3, +IPP	+3, +IPP
leren 'learn, teach'	+3, +IPP	–3, +IPP	+3, +IPP

Table 1.2: Third construction verbs

Table 1.2 shows that most verbs are only mentioned by two of the authors. With respect to the verbs at the bottom there is mainly disagreement with regard to their IPP status. Only for *helpen* 'help' and *leren* 'learn, teach' Klooster (2001: 254–255) mentions that they cannot occur in the third construction, whereas Rutten (1991) and IJbema (2002) consider them verbs of the third construction.

1.6 Cluster creeping

In a final set of constructions the IPP effect shows up (if the sentence is put in the perfect tense), but the verb cluster is interrupted with non-verbal material, as in (1.19).

Such sentences conflict with the principle that the verb cluster cannot be interrupted by non-verbal elements (cf. section 1.1).

(1.19) Ze hadden al vaker *proberen* duidelijk *te maken* dat er niets van
they had al more-often try clear to make that there nothing of
klopte.
agreed
'They frequently tried to make it clear that nothing was correct.'

According to Haeseryn et al. (1997), instances of *cluster creeping* occur more often in Belgian Dutch than in Netherlandic Dutch. Although normative grammars generally state that it is not allowed to interrupt a verb cluster by non-verbal elements, there are some exceptions. Haeseryn et al. (1997: 1355–1363) provide an extensive discussion of such constructions. A summary is given in the remainder of this section.

1.6.1 A typology of cluster creepers

Haeseryn et al. (1997) mention three types of *cluster creepers*.[5] The first type of cluster creepers consists of *inherent* parts of the verb phrase, such as predicative adjectives and non-verbal parts of idiomatic expressions. Those elements canonically appear just before the second pole, as in (1.20a), but they can also be included in the verb cluster, as in (1.20b) (Haeseryn et al. 1997: 1358). Note that (1.19) is also an example of this type.

(1.20) a. ...dat hij zich niet *bang* zal laten maken.
...that he himself not afraid will let make

b. ...dat hij zich niet zal laten *bang* maken.
...that he himself not will let afraid make
'...that he will not be frightened.'

A second category of cluster creepers consists of stranded adpositions, often as the second part of pronominal adverbs. Canonically those adpositions are realized before the verb cluster (1.21a), but they may also occur within the cluster (1.21b).

[5]Besides the interruption of verbal clusters, Haeseryn et al. (1997) mention the interruption of progressive *aan het*-infinitive constructions by non-verbal elements in this context as well. An example is *Vader is alweer aan het koffie zetten* (instead of *Vader is alweer koffie aan het zetten*) 'Father is making coffee again'. As *aan het*-infinitives never appear as a part of the verb cluster, they are not considered here.

(1.21) a. . . . dat hij daar nog *aan* moet denken.
. . . that he there still on must think

b. . . . dat hij daar nog moet *aan* denken.
. . . that he there still must on think
'. . . that he still needs to think about that.'

This type of adposition stranding within the cluster is considered typical of Belgian Dutch (Haeseryn et al. 1997: 1362).

A third type that is also typical of Belgian Dutch but less common than adposition stranding is cluster creeping by an object or an adverbial modifier (Haeseryn et al. 1997: 1362):

(1.22) a. . . . dat ik altijd al *Duits* wilde leren.
. . . that I always already German wanted learn

b. . . . dat ik altijd al wilde *Duits* leren.
. . . that I always already wanted German learn
'. . . that I have always wanted to learn German.'

Haegeman & van Riemsdijk (1986) discuss several constructions for West-Flemish, a regional variant of Dutch spoken in Belgium, such as (1.23a). Most speakers consider the corresponding construction in (Standard) Dutch ungrammatical (1.23b). What differentiates (1.23b) from (1.22b) is the presence of a determiner: While cluster creeping by bare nominals is more common, NPs with a determiner are rarely used in the verb cluster.

(1.23) a. WF . . . da Jan wilt *een hus* kopen.
. . . that Jan wants a house buy

b. DU * . . . dat Jan wil *een huis* kopen.
. . . that Jan wants a house buy
'. . . that Jan wants to buy a house.'

Besides *genuine* cases of cluster creeping, Haeseryn et al. (1997) mention several constructions that look like cluster creeping but should not be treated as such. For example, separable verb particles (SVPs) are not considered as cluster creepers if they occur within the verb cluster. They argue that in the case of SVPs, constructions in which the SVP is realised in front of the verb cluster (1.24a) are less preferred than constructions in which the SVP is realised within the cluster (in front of the main verb or as a part of it), as in (1.24b) (Haeseryn et al. 1997: 1357–1358).[6]

[6]Haeseryn et al. (1997) consider constructions like (1.24a) typical of spoken (Netherlandic) Dutch.

(1.24) a. . . . dat hij haar *op* moet bellen.
. . . that he her up must call

b. . . . dat hij haar moet *op*bellen.
. . . that he her must up-call
'. . . that he must call her.'

The fact that Haeseryn et al. (1997) do not treat SVPs as real cluster creepers as opposed to *inherent* parts of the verb phrase leads to classification problems, since the distinction between SVPs and inherent parts is often hard to draw (Haeseryn et al. 1997: 1359). Consider for example *koffiedrinken* 'drink coffee' versus *champagne drinken* 'drink champagne'. Are those examples separable verbs or regular combinations of a verb and a noun? In order to avoid this uncertainty, both SVPs and inherent parts of the verb phrase will be treated as cluster creepers, which is in line with amongst others Evers (2003), Wurmbrand (2005), and Broekhuis & Corver (2015).

Wurmbrand (2005) points out that there is a hierarchy of elements that may occur as cluster creepers. It is presented in Table 1.3.

Separable particles	Adverbs[7] Idioms Bare nouns	Indefinite objects PPs	Definite objects

Table 1.3: Hierarchy of cluster creepers (Wurmbrand 2005: 275)

If a language allows cluster creeping of a certain category, it also allows cluster creeping of the categories that occur more to the left in the hierarchy. For instance, if a language allows bare nouns in the verb cluster, it also allows for separable verb particles within the cluster. Wurmbrand states that adverbs, idiom chunks, and bare nouns marginally occur in Dutch verb clusters, but that separable verb particles are more often realized within the cluster (Wurmbrand 2005: 275). In the corpus study in chapter 5 it will be investigated to what extent Wurmbrand's hierarchy is visible in the data.

[7]The adverbs in Wurmbrand's hierarchy are the adverbs that canonically occur in front of the verb (cluster), i.e. the adverbs that are known as *low adverbs* in transformational grammar, in contrast to sentence adverbs or *high adverbs*, following Cinque (1999).

1.6.2 Position of the cluster creepers

Cluster creeping is only possible if the main verb does not occur at the front of the cluster, since the non-verbal element cannot occur after the main verb, as shown in (1.25).

(1.25) a. * . . . dat hij gedronken *koffie* heeft.
. . . that he drunk coffee has
Intended: '. . . that he has drunk coffee.'

b. * . . . dat hij drinken *koffie* wil.
. . . that he drink coffee wants
Intended: '. . . that he wants to drink coffee.'

Therefore, cluster creeping occurs more often in infinitival constructions than in constructions with a participle, since infinitives are usually realised at the end of the verb cluster, as opposed to participles (Haeseryn et al. 1997: 1355–1356). See also Hoekstra (2010) and Salzmann (2013) for a discussion on the relation between verb order within the cluster and cluster creeping.

The canonical position of a cluster creeper is just before the main verb, but in clusters with more than two verbs it may also occur more to the front of the verb cluster, as in (1.26) (Haeseryn et al. 1997: 1357).

(1.26) . . . dat hij haar had *op* moeten bellen.
. . . that he her had up must call
'. . . that he had to call her.'

In fact, the observation that SVPs can also occur inside the cluster without being adjacent to the main verb is an additional argument for treating them as genuine cluster creepers.

1.7 Word order variation

In section 1.2 it was already mentioned that the linear order of a verb cluster may deviate from the order of selection, as illustrated in (1.5), repeated in (1.27).

(1.27) a. . . . dat zij hem zou_1 $moeten_2$ $gezien_4$ $hebben_3$.
. . . that she him would must seen have

b. . . . dat zij hem zou_1 $gezien_4$ $moeten_2$ $hebben_3$.
. . . that she him would seen must have

c. ... dat zij hem gezien$_4$ zou$_1$ moeten$_2$ hebben$_3$.
... that she him seen would must have
'... that she should have seen him.'

Several variations on the linear order exist, depending on the regiolect or dialect and/or the type of elements that are used within the cluster. Word order variation in verb clusters does not influence the meaning of the sentence, in contrast to word order variation in other parts of the sentence. Changing the word order of constituents within the sentence usually causes a change in meaning, as shown in (1.28).

(1.28) a. Hij zag dat de hond de kat achtervolgde.
he saw that the dog the cat chased
'He saw that the dog chased the cat.'

b. Hij zag dat de kat de hond achtervolgde.
he saw that the cat the dog chased
'He saw that the cat chased the dog.'

Changing the word order of the elements within NPs also results in a change in meaning (1.29a–b), or in ungrammatical constructions (1.29c–d).

(1.29) a. die vijf mooie bloemen
those five beautiful flowers
'those five beautiful flowers'

b. die mooie vijf bloemen
those beautiful five flowers
'those beautiful five flowers'

c. * die vijf bloemen mooie
those five flowers beautiful

d. * die mooie bloemen vijf
those beautiful flowers five

For a discussion on word order variation within verb clusters versus variation within other parts of the sentence, see Barbiers (2008).

The literature on word order variation within verb clusters is vast. Especially the word order variation in Dutch two-verb clusters with a finite verb and a past participle has intrigued linguists for decades. This variation is also known as the red (auxiliary – past participle) versus the green word order (past participle – auxiliary), based on the colors used for the dialect maps in Pauwels (1953). It is often argued that the red and the green order are almost in free variation.

(1.30) a. . . . dat hij een pakje *heeft gebracht*. (red order)
. . . that he a package has delivered

b. . . . dat hij een pakje *gebracht heeft*. (green order)
. . . that he a package delivered has
'. . . that he has delivered a package.'

Besides descriptive and theoretical work on the topic, such as Haeseryn et al. (1997), there are diachronic accounts investigating the origin of the variation. Coussé (2006, 2008), for instance, puts the synchronic word order variation in a historical perspective. Quantitative investigations to model the choice between the two word orders are conducted in De Sutter (2005, 2009), Arfs (2007), and Bloem et al. (2014).

Wurmbrand (2004, 2005) takes Dutch into account in her cross-linguistic study of West Germanic verb clusters, based on the literature. Apart from (Standard) Dutch, she includes Afrikaans, Frisian, Standard German and West Flemish in her overview.[8] She states that in two-verb clusters the 1-2 and the 2-1 order are possible in auxiliary-participle constructions, cf. (1.30). In modal-infinitive combinations the 2-1 order is only possible if the modal is a finite verb (e.g. *lezen kan* 'read.INF can.FIN'). Regarding the word order in Dutch three-verb clusters Wurmbrand points out that the canonical 1-2-3 order is obligatory in clusters with a finite verb and two infinitives. If the cluster contains a past participle, the participle can obtain any position within the cluster. In Wurmbrand's studies clusters with *te*-infinitives are not taken into account.

Barbiers (2005) describes the results of the SAND project (Syntactic Atlas of the Dutch Dialects) regarding Dutch three-verb clusters.[9] Within the SAND-project, word order variation in verb clusters in the entire Dutch language area was investigated by means of oral interviews (following a pilot study based on questionnaires). The results show that five out of the six logical word orders of three-verb clusters actually occur in the different varieties of Dutch. Depending on the region and on the type of verb cluster, the use of certain orders is excluded. The results for the different types of clusters are summarised in Table 1.4 (Barbiers 2005: 239). The constructions under investigation are all verb-final constructions, so the finite verb is included in the verb clusters. Moreover, constructions containing a *te*-infinitive were not considered.

[8] Besides the cross-linguistic overview, Wurmbrand (2004) furthermore makes use of questionnaires to collect data on dialect variation with respect to the word order of German verb clusters.

[9] For a description of the SAND project, see `http://www.meertens.knaw.nl/projecten/sand/sandeng.html`.

TYPE OF CLUSTER	WORD ORDER
$Modal_1$ $Modal_2$ $Verb_3$	1-2-3
e.g. *moet kunnen werken*	1-3-2
'must can work'	* 2-1-3
	* 2-3-1
	3-1-2
	3-2-1
$Modal_1$ $Auxiliary_2$ $Verb_3$	1-2-3
e.g. *moet hebben gemaakt*	1-3-2
lit. 'must have made'	* 2-1-3
	* 2-3-1
	3-1-2
	3-2-1
$Auxiliary_1$ $Aspectual/Modal_2$ $Verb_3$	1-2-3
e.g. *is gaan zwemmen*	1-3-2
lit. 'is go swim'/	* 2-1-3
heeft kunnen zwemmen	2-3-1
lit. 'has can swim'	* 3-1-2
	3-2-1

Table 1.4: Word order variation in Dutch three-verb clusters (Barbiers 2005)

Table 1.4 shows that the word order 2-1-3 is excluded in all cluster types. Furthermore, their is a remarkable difference in ordering possibilities between the IPP constructions (Aux–Asp/Mod–V) and the constructions that have a modal as the hierarchically highest verb in the cluster. In several dialects, the word order 2-3-1 (e.g. *kunnen zwemmen heeft* 'can swim has') is allowed in IPP constructions, whereas this word order is excluded in the other cluster types. In contrast, the 3-1-2 order is excluded in IPP constructions, but not in the other cluster types (e.g. *gemaakt moet hebben* 'made must have').

In section 1.4 it was suggested that the difference between IPP constructions and the third construction is the form of the verb selected by the auxiliary of the perfect, as illustrated in example (1.17), repeated in (1.31).

(1.31) a. Ze hadden al vaker duidelijk *geprobeerd* te maken dat er
they had already more-often clear tried to make that there
niets van klopte.
nothing of agreed

b. Ze hadden al vaker duidelijk *proberen* te maken dat er
they had already more-often clear try to make that there
niets van klopte.
nothing of agreed
'They had frequently tried to make it clear that nothing was correct.'

If one has a closer look at the differences between IPP constructions and the third construction in verb-final sentences it turns out that both constructions allow a different word order. While the findings in Table 1.4 illustrate that none of the three-verb clusters shows the 2-1-3 order, this is exactly what one finds in the third construction:

(1.32) a. ...dat Jan Marie heeft$_1$ geprobeerd$_2$ te kussen$_3$.
...that Jan Marie has tried to kiss

b. ...dat Jan Marie geprobeerd$_2$ heeft$_1$ te kussen$_3$.
...that Jan Marie tried has to kiss

c. * ...dat Jan Marie geprobeerd$_2$ te kussen$_3$ heeft$_1$.
...that Jan Marie tried to kiss has
'...that Jan has tried to kiss Marie.'

As observed by den Besten & Rutten (1989: 46–49), in the third construction the finite verb can immediately precede (1.32a) or follow (1.32b) the past participle. Sentences like (1.32c) are not attested in the literature but are probably ungrammatical for most speakers.[10] The word order variation in (1.32) suggests that in those constructions the auxiliary and the past participle form a cluster, while the *te*-infinitive is in the Nachfeld.

In contrast, in IPP constructions with a *te*-infinitive the auxiliary of the perfect has to be either in the front or at the end of the verb cluster:

(1.33) a. ...dat Jan Marie heeft$_1$ proberen$_2$ te kussen$_3$.
...that Jan Marie has try to kiss

b. ...dat Jan Marie proberen$_2$ te kussen$_3$ heeft$_1$.
...that Jan Marie try to kiss has

[10] A few examples could be found via a Google search though, e.g. *Ik ontvang nog steeds smsjes nadat men mij geprobeerd te bellen heeft* 'I still receive SMSes after one has tried to call me' [Google.be, 14-08-2014].

c. * . . . dat Jan Marie proberen$_2$ heeft$_1$ te kussen$_3$.
. . . that Jan Marie try has to kiss
'. . . that Jan has tried to kiss Marie.'

The examples in (1.33) show word order variation that is in line with the findings of Barbiers (2005) for IPP constructions without a *te*-infinitive (cf. Table 1.4). While the ascending 1-2-3 order is preferred over the 2-3-1 order in most varieties of Dutch, the order in (1.33c) is ungrammatical in all varieties of Dutch. This suggests that in IPP constructions the *te*-infinitive is in the second pole, rather than in the Nachfeld.

In contrast to Wurmbrand (2004, 2005) and Barbiers (2005), who discuss word order variation in all Dutch dialects, Broekhuis & Corver (2015: 1097) focus more on variation within Standard (and colloquial) Dutch. They formulate the following generalizations with respect to word order in Dutch verb clusters:

Generalization I: Past/passive participles either precede or follow their governing auxiliary.
Generalization II: *Te*-infinitives follow their governing verb.
Generalization III: Bare infinitives follow their governing verb.

With respect to generalization I, example (1.27b) shows that participles do not have to immediately precede the governing verb. While generalization II holds for all *te*-infinitives selected in the cluster, Broekhuis & Corver (2015) indicate that generalization III might be too restrictive for two-verb clusters. In Standard Dutch certain bare infinitives can occur before their selector in two-verb cluster, but not in longer clusters:

(1.34) a. . . . dat hij het nog altijd niet begrijpen$_2$ kan$_1$.
. . . that he it still always not understand can
'. . . that he still cannot understand it.'

b. * . . . dat hij het nog altijd niet hebben$_2$ begrepen$_3$ kan$_1$
. . . that he it still always not have understand can
Intended: '. . . that he still has not been able to understand it.'

It is generally assumed that the 2-1 order is grammatical for constructions in which the finite verb is a modal, but there is some disagreement in the literature whether it is also allowed for other verb classes, such as aspectuals, perception verbs, and causatives. For a discussion, see Broekhuis & Corver (2015: 1096). Moreover, Broekhuis & Corver (2015) are inconclusive whether such constructions should be considered

as part of Standard Dutch, or rather as a regional and/or stilistic variant. The corpus study presented in chapter 5 will show whether the generalizations hold for the constructions encountered in the treebanks.

1.8 Verbal complement types

Another criterion that is often linked to the difference between clustering and non-clustering constructions, is the nature of the verbal complement. Klooster (2001) links up the IPP effect with the type of verbal complement. He states that verbs selecting a bare infinitive obligatorily appear as IPP in the perfect (1.35a), whereas verbs selecting a *te*-infinitive optionally appear as IPP (1.35b), or cannot appear as IPP (1.35c).

(1.35) a. . . . dat ik hen heb horen aankomen.
. . . that I them have hear arrive
'. . . that I've heard them coming.'

b. Hij heeft dat trachten/getracht te maken.
he has that try/tried to make
'He has tried to fix that.'

c. Hij had beloofd/*beloven te komen.
he had promised/*promise to come
'He had promised to come.'

However, there is a small group of verbs that select a *te*-infinitive but obligatorily appear as IPP, such as *(be)horen* 'ought to', *dienen* 'have to', and *zitten* 'sit' (Klooster 2001: 242).

(1.36) Hij heeft de hele tijd zitten/*gezeten te slapen.
he has the whole time sit/*sat to sleep
'He was sleeping all the time.'

Some speakers can drop the *te* in those constructions, without changing the meaning of the sentence. In constructions with a past participle, *te* deletion is ungrammatical:

(1.37) Hij heeft de hele tijd zitten slapen.
he has the whole time sit sleep
'He was sleeping all the time.'

(1.38) * Hij heeft dat getracht maken.
he has that tried make

It is furthermore possible to divide the set of verbs selecting a *te*-infinitive into constructions that can be preceded by the complementizer *om* 'for', and constructions that do not allow *om*. Klooster points out that the appearance of *om* is excluded in constructions where the IPP effect obligatorily shows up (1.39a). The optional IPP verbs can occur with *om* (1.39b), but not in the third construction or if the IPP effect is present (1.39c) (Klooster 2001: 253).

(1.39) a. * Hij heeft de hele tijd zitten om te slapen.
he has the whole time sit for to sleep
'Intended: He was sleeping all the time.'

b. Hij heeft getracht om dat te maken.
he has tried for that to make
'He has tried to fix that.'

c. * Hij heeft dat trachten/getracht om te maken.
he has that try/tried for to make

The examples in (1.39) show that if *om* is present, cluster formation is excluded.

Klooster (2001: 246–247) provides an overview of verbs selecting a non-finite complement. Although the list is not as extensive as the list of verbs provided by Haeseryn et al. (1997), it is more interesting from a linguistic point of view since the verbs are categorized along different (linguistically relevant) criteria, such as IPP, the occurrence of *om*, and the type of verbal complement they select.

1.9 Cross-serial dependencies

Because of the pole structure and the quasi-impenetrability of the verb cluster the verbs in the cluster are usually not adjacent to their arguments. In example (1.40) *ik* 'I' is separated from *horen* 'hear' and the arguments *de boer* 'the farmer' and *een varken* 'a pig' are separated from *slachten* 'slaughter'. Since non-verbal elements generally cannot occur within the verb cluster, they have to be realised elsewhere; usually they end up in the Mittelfeld.[11]

[11] Haeseryn et al. (1997) mainly describe the internal structure of the verb cluster. Less attention is devoted to the position of the arguments vis-à-vis the verbs in the cluster. Instead, the position of the arguments vis-à-vis adjuncts and other arguments in the Mittelfeld is extensively discussed (e.g. direct object versus indirect object).

(1.40) ...dat ik$_A$ de boer$_B$ een varken$_B$ hoorde$_A$ slachten$_B$.
...that I the farmer a pig heard slaughter
'...that I heard the farmer slaughter a pig.'

In constructions like (1.40) with a canonical (ascending) verb cluster and multiple non-verbal arguments the word order of the arguments is parallel to the word order of the verbs in the cluster. The order of arguments vis-à-vis their heads is not random: In Dutch, sentences with verb clusters typically show an ABAB pattern of dependencies, hence the name *cross-serial dependencies*.[12]

1.10 The boundaries of the second pole

In section 1.2 a verb cluster was defined as 'a sequence of two or more verbs in the second pole'. It was furthermore shown that the linear order of the verbs may deviate from the order of selection, and that clustering verbs typically appear as IPP if they are put in the perfect tense. Despite those criteria, it is not always easy to distinguish clustering verbs – or rather, the clustering use of verbs – from verbs that select their complement in the Mittelfeld or the Nachfeld. The latter will be discussed in section 1.10.1, whereas the former will be discussed in section 1.10.2.

1.10.1 In the second pole or in the Nachfeld?

Constructions with *te*-infinitives are tricky, since *te*-infinitives may be selected in the cluster or in the Nachfeld if they are selected as a verbal complement.[13] The former is invariably the case if its selector is an IPP verb, which was motivated in section 1.7. If the *te*-infinitive is selected by a past participle, it is in the Nachfeld.

As a consequence, constructions in which an optionally IPP verb appears in the finite form are ambiguous between clustering and non-clustering constructions:

(1.41) a. dat Jonas gitaar probeert te spelen.
that Jonas guitar tries to play

[12]Note that in example (1.40) the object of *horen* 'hear' is not *de boer* 'the farmer', but the embedded verb and its complements (*de boer het varken slachten*). Therefore, *de boer* is labelled B and not A. The semantics of the sentence provides evidence for that, as it is the pig that is heard by the subject and not the farmer.

[13]*Te*-infinitives can also appear in the Mittelfeld if they are predicative, cf. section 1.10.2.

b. dat Jonas probeert gitaar te spelen.
that Jonas tries guitar to play
'that Jonas tries to play the guitar.'

Examples like (1.41a) are canonically treated as clustering constructions, whereas (1.41b) is usually treated as a construction without a verb cluster. If the constructions in (1.41) are put into the perfect tense, however, we get the following possibilities:

(1.42) a. dat Jonas gitaar heeft proberen te spelen.
that Jonas guitar has try.IPP to play

b. dat Jonas gitaar heeft geprobeerd te spelen.
that Jonas guitar has try.PSP to play

c. dat Jonas heeft proberen gitaar te spelen.
that Jonas has try.IPP guitar to play

d. dat Jonas heeft geprobeerd gitaar te spelen.
that Jonas has try.PSP guitar to play
'that Jonas has tried to play the guitar.'

If the construction in (1.41a) is put into the perfect, it can correspond with either the clustering IPP construction (1.42a) or the third construction (1.42b). (1.41b) corresponds to either a construction with an interrupted cluster (1.42c) or a construction in which all arguments of *geprobeerd* 'tried' are selected in the Nachfeld.

In the corpus study in chapter 5 it will be investigated in further detail whether examples like (1.41) are indeed ambiguous between clustering and non-clustering constructions, or whether the corpus data suggest an analysis in favour of a clustering or non-clustering construction.

1.10.2 In the second pole or in the Mittelfeld?

Predicative complements

The constructions containing *te*-infinitives considered so far showed that *te*-infinitives typically appear at the end of the cluster or in the Nachfeld. However, there are also cases in which the *te*-infinitive canonically precedes the finite verb:

(1.43) a. ...dat het wel te doen is.
...that it rather to do is
'...that it can be done.'

b. ? . . . dat het wel is te doen.
. . . that it rather is to do

Van Eynde (2015) argues that the *te*-infinitives in constructions like (1.43) behave as predicative complements, rather than as the infinitival complement of a clustering verb. The canonical position of predicative complements in Dutch is the right part of the Mittelfeld, whereas the canonical position of a *te*-infinitive that is selected as a verbal complement is at the end of the verb cluster or in the Nachfeld. Besides the difference in word order, Van Eynde (2015) provides several criteria to distinguish predicative infinitives from verbal complements. Predicative infinitives have a passive and modal interpretation, as exemplified by the translation of (1.43a). Furthermore, their subject must be referential (1.44a), they can take a degree marker (1.44b), and they can also occur in attributive position (1.44c).

(1.44) a. * . . . dat het wel te sneeuwen is.
. . . that it rather to snow is

b. . . . dat het zeer te mijden is.
. . . that it very to avoid is
'. . . that it is very much to be avoided.'

c. een te mijden buurt
a to avoid neighbourhood
'a neighbourhood to be avoided'

Furthermore, Van Eynde (2015) points out that the predicative *te*-infinitives behave similarly to present participles:

(1.45) a. Wat er ook gaande is.
what there also going is
'Whatever there is going on.'

b. * Wat er ook is gaande.
what there also is going

A present participle such as *gaande* 'going' in (1.45) is never a part of the verb cluster. Such participles canonically occupy a position in the Mittelfeld, similar to predicative *te*-infinitives; they cannot occur after their selector.

Passive participles can also be used as predicative complements. An example is *gewend* 'used' in (1.46).

(1.46) a. dat hij er aan gewend geraakt.
that he there to used becomes
'that he gets used to it'

b. * dat hij er aan geraakt gewend.
that he there to becomes used

Also in (1.46), the predicative participle cannot occur after its selector, in contrast to participles selected by clustering verbs, indicating that it also obtains a position in the Mittelfeld.

In constructions where two participles follow each other, only one participle can be part of the cluster. For example, in (1.47) *gewend* 'used' is in the Mittelfeld, while *geraakt is* 'become is' occupies the second pole. This is confirmed by the word order variation:

(1.47) a. dat hij er intussen aan gewend geraakt is.
that he there meanwhile to used become is

b. dat hij er intussen aan gewend is geraakt.
that he there meanwhile to used is become
'that he got used to it in the meantime'

Even though predicative participles like *gewend* 'used' are canonically placed at the end of the Mittelfeld and not in the second pole, cases of cluster creeping are not uncommon.

(1.48) a. dat hij er intussen aan is gewend geraakt.
that he there meanwhile to is used become
'that he got used to it in the meantime'

b. * dat hij er intussen aan geraakt gewend is.
that he there meanwhile to become used is

(1.48) illustrates that predicative participles can appear as cluster creepers, but only if they occupy a position before their selector. In this respect they are like predicative adjectives (see also section 1.6).

Fixed expressions

Besides predicative *te*-infinitives there is a set of fixed expressions in which the *te*-infinitive canonically precedes its selector. The examples in (1.49) show that the verb *komen* 'come' typically follows the *te*-infinitive if it is used in the expression *te weten komen* 'find out' (1.49a). If it is used as a clustering verb, it canonically precedes that complement (1.49b).

(1.49) a. . . . dat hij het niet te weten komt / ?komt te weten.
. . . that he it not to know comes comes to know
'. . . that he does not find it out.'

b. . . . als het daar niet komt te liggen / *te liggen komt.
. . . if it there not comes to lie to lie come
'. . . if it will not end up lying there.'

The different uses of the verb *komen* in (1.49) are even more apparent if the sentences are put in the perfect tense:

(1.50) a. . . . dat hij het niet te weten is gekomen/*komen.
. . . that he it not to know is come.PSP/come.IPP
'. . . that he has not found it out.'

b. . . . als het daar niet is komen/*gekomen te liggen.
. . . if it there not is come.IPP/come.PSP to lie
'. . . if it has not ended up lying there.'

The lack of the IPP effect in (1.50a) indicates that *komen* is not used as a clustering verb, but it definitely allows clustering in (1.50b), where *komen* obligatorily appears as IPP.

Similar to predicative complements, parts of fixed expressions are typical cluster creepers:

(1.51) . . . dat hij het niet is *te weten* gekomen.
. . . that he it not is to know come.PSP
'. . . that has not found it out.'

Expressions of the type *te weten komen* 'find out' or *te wachten staan* 'await' behave similarly to fixed expressions and idiom chunks with non-verbal elements.

(1.52) a. . . . dat hij met dat probleem *rekening* moet houden.
. . . that he with that problem account has hold

b. . . . dat hij met dat probleem moet *rekening* houden.
. . . that he with that problem has account hold
'. . . that he has to take into account that problem.'

As shown in (1.52), parts of fixed expressions canonically appear in the Mittelfeld, but they are also typical instances of cluster creepers (see also section 1.6).

1.11 Cluster formation in German

In this section the phenomena related to cluster formation that were described for Dutch in the previous sections will be sketched for German. Although the main focus of this thesis will be on Dutch, the theoretical literature on German verb clusters is very extensive. Since the canonical treatment of Dutch verb clusters within HPSG is based on the analysis of German verb clusters (see chapter 3), it is relevant to have a look at the phenomena in the descriptive literature as well.

1.11.1 German sentence structure

As is the case for Dutch sentence structure, German sentence structure is characterised by two poles or sentence brackets that serve as orientation points for the other elements in the sentence. For an introduction on the topological field model in German, see amongst others Höhle (1983) and Duden (2006: 874–901).

In verb-initial sentences, the finite verb occupies the first pole (1.53a–b). If the sentence contains non-finite verb forms, they occur in the second pole (1.53c). In verb-final sentences a complementizer is placed in the first pole, while all verbs form a *Verbalkomplex* or cluster in the second pole (1.53d) (Duden 2006: 480).

(1.53) a. Ich *sehe* dich.
I see you
'I see you.'

b. *Siehst* du mich?
see you me
'Do you see me?'

c. Ich *habe* dich nicht *gesehen*.
I have you not seen
'I haven't seen you.'

d. . . . , *dass* ich dich nicht *gesehen habe*.
. . . , that I you not seen have
'. . . that I didn't see you.'

Like Dutch verb clusters, German verb clusters generally cannot be split up by non-verbal elements, as illustrated in (1.54) (Duden 2006: 480–481).

(1.54) a. . . . , weil man nicht alles *gelesen haben muss*.
. . . , because one not everything read.PSP have.INF must.FIN

b. * . . . , weil man nicht *gelesen* alles *haben* *muss*.
. . . , because one not read.PSP everything have.INF must.FIN

c. * . . . , weil man nicht *gelesen* *haben* alles *muss*.
. . . , because one not read.PSP have.INF everything must.FIN
'. . . because one need not have read everything.'

(1.54a) shows that the canonical verb order in German verb clusters is the mirror image of the order in the corresponding Dutch constructions, since verbs selecting a verbal complement in the second pole canonically occur to the right of their complement, cf. section 1.11.4.

1.11.2 Coherent versus incoherent structures

Following Bech (1955) the literature on German syntax distinguishes *coherent* from *incoherent* constructions. The distinction between coherent and incoherent constructions is similar to the distinction between clustering and non-clustering constructions in Dutch: Verbs occurring in coherent constructions form a cluster and select their non-verbal complements in the Mittelfeld (1.55a), whereas verbs in incoherent constructions select their complements in the Nachfeld (1.55b).

(1.55) a. . . . , weil Lisa ein bisschen *zu schlafen versucht*.
. . . , because Lisa a little to sleep tries

b. . . . , weil Lisa *versucht* ein bisschen *zu schlafen*.
. . . , because Lisa tries a little to sleep
'. . . because Lisa tries to sleep a little.'

Optionally coherent verbs, such as *versuchen* 'try' in (1.55) can occur in clustering and non-clustering constructions, whereas obligatorily coherent verbs such as *scheinen* 'appear' in (1.56) invariably occur in clustering constructions. Constructions in which they select their complements in the Nachfeld are ungrammatical.

(1.56) a. . . . , weil Karl das Buch zu lesen *scheint*.
. . . , because Karl the book to read seems
'. . . because Karl seems to be reading the book.'

b. * . . . , weil Karl *scheint* das Buch zu lesen.
. . . , because Karl seems the book to read

The previous examples illustrate that *zu*-infinitives can be selected in the second pole or in the Nachfeld. In contrast, bare infinitives (1.57) and participles (1.58) are always selected in the second pole (Müller 2002: 43).

(1.57) a. dass Karl das Buch lesen wird.
that Karl the book read.INF will.FIN
'that Karl will read the book.'

b. * dass Karl wird das Buch lesen.
that Karl will.FIN the book read.INF

(1.58) a. dass Karl das Buch gelesen hat.
that Karl the book read.PSP has
'that Karl has read the book.'

b. * dass Karl hat das Buch gelesen.
that Karl has the book read.PSP

According to Müller (2002: 67) there are no obligatorily incoherent verbs. Dutch and German differ on this point, as Dutch has a set of verbs selecting a verbal complement that are not clustering (e.g. *merken* 'notice'), cf. section 1.3.

1.11.3 Infinitivus Pro Participio

In German the IPP effect or *Ersatzinfinitiv* occurs in similar environments as in Dutch: A verb selected by an auxiliary of the perfect appears as an infinitive instead of a past participle. The number of verbs in which the IPP effect can show up is smaller compared to Dutch. Some verbs are optionally IPP verbs in German while their Dutch cognates are obligatory IPP verbs. Duden (2006) lists the verbs in Table 1.5 as IPP verbs.

OBLIGATORY IPP VERBS		OPTIONAL IPP VERBS	
brauchen	'need'	*sehen*	'see'
dürfen	'be allowed to'	*hören*	'hear'
können	'can'	*fühlen*	'feel'
mögen	'like'	*spüren*	'notice'
müssen	'must, have to'	*lassen*	'let'
sollen	'have to, will'	*helfen*	'help'
wollen	'want'	*lehren*	'teach'
		lernen	'learn'

Table 1.5: German IPP verbs (Duden 2006: 473)

The modal verbs always appear as IPP if they select another verb, as shown in (1.59) (Duden 2006: 473).

(1.59) a. Wir haben nichts tun *müssen*.
we have.FIN nothing do.INF must.IPP
'We didn't have to do anything.'

b. * Wir haben nichts tun *gemusst*.
we have.FIN nothing do.INF must.PSP

If they occur as main verbs in the perfect tense, those verbs generally occur as a past participle, such as *gekonnt* in (1.60a). However, some southern German dialects show IPP in those constructions as well (1.60b) Duden (2006: 473). This kind of IPP does not occur in Dutch.[14]

(1.60) a. Das hat niemand *gekonnt*.
that had no one can.PSP
'No one was able to do that'

b. Ich hätte es bestimmt nicht *können*.
I would have it definitely not can.IPP
'I definitely would not have been able to do that.'

Another set of verbs that may appear as IPP are the perception verbs *sehen* 'see', *hören* 'hear', *fühlen* 'feel', and *spüren* 'notice', as well as the causative *lassen* 'let'. They are optional IPP verbs, as constructions with a past participle are grammatical as well (1.61).

(1.61) Die Familie hatte das Ende offenbar *kommen sehen/gesehen*.
the family have.FIN the end apparently come.IPP see.INF/see.PSP
'Apparently the family had noticed the end was coming.'

The following verbs can appear as IPP if they select a bare infinitive (without *zu*): *helfen* 'help', *lehren* 'teach' and *lernen* 'learn'. Although IPP is not excluded, these verbs occur more frequently in constructions with a past participle (Duden 2006: 473).

In section 1.11.2 it was suggested that the coherent and incoherent verbs in German correspond to clustering and non-clustering constructions in Dutch. The criteria to distinguish clustering constructions in both languages differ, however. The IPP effect in Dutch can be used as a diagnostic for clustering, but in German it cannot be used in the same way. If a verb occurs as IPP, it is used as a clustering verb, but verbs that cannot trigger IPP can be clustering nonetheless. An example is *versuchen* 'try' in (1.62).

[14] Plank (2000) notes that in Bavarian the participle forms can be used instead of the substitute infinitive in constructions like (1.60b); it is not the case that the form of the past participle and the infinitive is similar in those dialects.

(1.62) weil Lisa ein bisschen zu schlafen versucht/*versuchen hat.
because Lisa a little to sleep.INF try.PSP/try.INF have.FIN
'because Lisa has tried to sleep a little.'

Verbs that are used as a clustering verb select both their verbal and their non-verbal complements to the left. In non-clustering constructions they select their verbal complement and (some of the) non-verbal arguments to the right, cf. the contrastive examples in section 1.11.2.

1.11.4 Word order variation

In (Standard) German the word order within the verb cluster is less free compared to Dutch. In verb-final constructions with two verbs in the cluster the auxiliary of the perfect always has to follow the main verb (1.63).

(1.63) a. ..., dass ich dich nicht gesehen$_2$ habe$_1$.
..., that I you not see.PSP have.FIN

b. * ..., dass ich dich nicht habe$_1$ gesehen$_2$.
..., that I you not have.FIN see.PSP
'... that I didn't see you.'

Similar to Dutch clustering verbs, German clustering verbs may select another clustering verb. As opposed to Dutch, each selecting verb canonically occurs to the right of its verbal complement in clusters with more than two verbs, as in example (1.64).

(1.64) weil man nicht alles gelesen$_3$ haben$_2$ muss$_1$
because one not everything read.PSP have.INF must.FIN
'because one does not need to have read everything.'

As was already discussed in sections 1.11.2 and 1.11.3, in German it can be determined by means of the word order alone whether the main verb is selected in the cluster or not, since verbal complements in the cluster are canonically selected to the left.

However, there are some constructions in which German deviates from the rigid descending order in verb clusters. If the IPP effect shows up, the auxiliary may appear at the beginning of the cluster. That phenomenon is also known as *Oberfeldumstellung* or *auxiliary flip* (Hinrichs & Nakazawa 1989). If the auxiliary of the perfect selects a modal verb, the Oberfeldumstellung is obligatory (1.65).

(1.65) a. . . . , was sie dennoch nicht hätte$_1$ sagen$_3$ sollen$_2$
. . . , what she still-not not would have say.INF should.IPP
'. . . what she still not should have said'

b. * . . . , was sie dennoch nicht sagen$_3$ sollen$_2$ hätte$_1$
. . . , what she still-not not say.INF should.IPP would have

Besides the 'flipped' order, the canonical descending order is grammatical as well in IPP constructions with the causative *lassen* and the perception verbs:

(1.66) a. . . . , warum Sie mich haben$_1$ kommen$_3$ lassen$_2$ . . . , als ich die Vögel
. . . , why you me have.FIN come.INF let.IPP . . . , when I the birds
habe$_1$ zwitschern$_3$ hören$_2$
have.FIN twitter.INF hear.IPP

b. . . . , warum Sie mich kommen$_3$ lassen$_2$ haben$_1$ . . . , als ich die Vögel
. . . , why you me come.INF let.IPP have.FIN . . . , when I the birds
zwitschern$_3$ hören$_2$ habe$_1$
twitter.INF hear.IPP have.FIN
'. . . why did you let me come . . . , when I heard the birds twitter'

By contrast, if an optional IPP verb occurs as a past participle, only the regular verb order can be used, as illustrated in (1.67) (Duden 2006: 480–481).

(1.67) a. . . . , warum Sie mich kommen$_3$ gelassen$_2$ haben$_1$ . . . , als ich die Vögel
. . . , why you me come.INF let.PSP have.FIN . . . , when I the birds
zwitschern$_3$ gehört$_2$ habe$_1$
twitter.INF hear.PSP have.FIN
'. . . why did you let me come . . . , when I heard the birds twitter'

b. * . . . , warum Sie mich haben$_1$ kommen$_3$ gelassen$_2$ . . . , als ich die Vögel
. . . , why you me have.FIN come.INF let.PSP . . . , when I the birds
habe$_1$ zwitschern$_3$ gehört$_2$
have.FIN twitter.INF hear.PSP

Note that the Oberfeldumstellung is not characteristic for IPP constructions alone, as it may also occur in constructions with the future auxiliary *werden* 'will'. Also in this case, it is optional (Duden 2006: 482).

(1.68) a. . . . , dass sie bald wird$_1$ ausziehen$_3$ müssen$_2$
. . . , that she soon will.FIN move out.INF must.IPP

b. . . . , dass sie bald ausziehen$_3$ müssen$_2$ wird$_1$
. . . , that she soon move out.INF must.IPP will.FIN
'that she will have to move out soon'.

The examples with IPP in (1.65) and (1.66) indicate that IPP constructions in German differ from Dutch IPP constructions in several respects. First, IPP in German is more closely linked to word order: In most cases, it is not allowed in strictly descending (e.g. 3-2-1) word orders, while this is the canonical word order in other verb clusters. Second, although the phenomenon occurs in German and Dutch, IPP in German is not as widespread as in Dutch: Table 1.5 contains fewer IPP verbs compared to Table 1.1, and there are fewer verbs in German that obligatorily appear as IPP. For a comparison between Dutch and German verb clusters, see den Besten & Edmondson (1983).

Although in general the German word order within verb clusters is more rigid compared to Dutch, the German dialects show more variation. For instance, some German dialects allow the *Zwischenstellung* besides the Oberfeldumstellung. In those cases, the auxiliary does not occur at the front of the verb cluster (1.69) (Wurmbrand 2004: 50).

(1.69) weil er ihr *helfen*$_3$ *hätte*$_1$ *müssen*$_2$
since he her help.INF have.FIN must.IPP
'. . . since he had to help her.'

The dialects spoken in Switzerland deviate the most from Standard German with respect to verb order. Zürich German, for instance, allows the strictly ascending order (i.e. the canonical word order in Dutch) besides the canonical descending order. An example is given in (1.70), quoted from Haegeman & van Riemsdijk (1986: 432).

(1.70) das er sini chind mediziin *wil*$_1$ *laa*$_2$ *studiere*$_3$.
that he his children medicine want.FIN let.IPP study.INF
'that he wants to let his children study medicine'.

For more details on verb clusters in non-standard varieties of German, see amongst others Wurmbrand (2004), and Bader & Schmid (2009).

1.11.5 The Third Construction and cluster creeping

As in Dutch, the third construction and constructions in which the verb cluster is interrupted by non-verbal material occur in German as well. Example (1.71a) is an illustration of the third construction in German. In contrast to clustering constructions (1.71b), the verbal complement *zu reparieren* 'to repair' is selected in the Nachfeld.

Similar to the clustering construction, but in contrast to the non-clustering construction in (1.71c), the non-verbal complement *es* 'it' occurs in the Mittelfeld (Kathol 2000: 243).

(1.71) a. dass er es versucht zu reparieren.
that he it tries to repair
'that he tries to repair it.'

b. dass er es zu reparieren versucht.
that he it to repair tries

c. dass er versucht es zu reparieren.
that he tries it to repair
'that he tries to repair it.'

Besides the third construction, instances of cluster creeping are also attested in German. An example is given in (1.72a), quoted from Kathol (2000: 242):

(1.72) a. dass er seiner Tochter *hätte* das Märchen *vorlesen dürfen*.
that he his daughter.DAT had the fairy tale.ACC read to may
'that he had been allowed to read the fairy tale to his daughter.'

b. dass er seiner Tochter das Märchen *hätte vorlesen dürfen*.
that he his daughter.DAT the fairy tale.ACC had read to may

In contrast to the corresponding clustering construction in (1.72b), the verb cluster in (1.72a) is interrupted by the NP *das Märchen* 'the fairy tale'. As in Dutch, such constructions are considered ungrammatical in Standard German.

It was discussed in section 1.6 that cluster creeping in Dutch is only possible if the main verb is not the first verb of the cluster. This generalization also holds for German, which implies that cluster creeping is only possible in the case of Obferfeldumstellung (1.72) or Zwischenstellung, cf. section 1.11.4.

1.11.6 Cross-serial dependencies

In German the non-verbal arguments of the verbs in the cluster generally occur in the Mittelfeld in an order that is similar to the Dutch word order (1.73).

(1.73) a. ..., als ich$_A$ die Vögel$_B$ *zwitschern*$_B$ *gehört*$_A$ *habe*$_A$
..., when I the birds twitter.INF hear.PSP have.FIN

b. ..., als ich$_A$ die Vögel$_B$ *habe*$_A$ *zwitschern*$_B$ *hören*$_A$
..., when I the birds have.FIN twitter.INF hear.IPP
'... when I've heard the birds twitter.'

Since the canonical order of verbs and their verbal complement is different in German and Dutch, the languages also differ in the way the non-verbal complements are organized vis-à-vis their verbal selectors. In constructions where Dutch uses cross-serial dependencies, showing an ABAB pattern, German makes use of embedded dependencies, revealing an ABBA pattern.

In (Standard) German the complements of the main verb are generally realized adjacent to the main verb, such as *die Vögel* and *zwitschern* in (1.73a). In the case of *Oberfeldumstellung* the complements are separated from the main verb by the auxiliary of the perfect (1.73b). If the infinitive-selecting verbs select complements of their own, like *hören* 'hear' in (1.73), they are separated by the main verb (1.73a) and sometimes by the main verb and the tense auxiliary (1.73b).

While in Standard German the canonical word order leads to embedded dependencies, the example of Zürich German given in (1.70) and repeated in (1.74) shows cross-serial dependencies:

(1.74) das er$_{A}$ sini chind$_{B}$ mediziin$_{B}$ *wil*$_{A}$ *laa*$_{A}$ *studiere*$_{B}$
that he his children medicine want.FIN let.IPP study.INF
'that he wants to let his children study medicine'.

The construction in (1.74) shows a similar pattern of dependencies as is common in Dutch sentences with verb clusters.

1.12 Conclusion

In this chapter an overview was given of how verb clusters and the phenomena related to cluster formation are treated in the descriptive literature. First a definition of the Dutch verb cluster was given, indicating that the order of selection in the cluster is not necessarily similar to the linear order of the verbs. In the literature it is canonically assumed that the occurrence of the IPP effect is an important indication for clustering. In addition, the word order within the cluster and the nature of the verbal complement provide information about the clustering abilities of a verb as well.

The occurrence of the third construction and constructions with cluster creepers indicates that Dutch verb clusters cannot be defined as just a contiguous sequence of verbs, as those constructions have different properties with respect to word order variation and the set of verbs that can participate in such constructions. Moreover, it is important to distinguish verb clusters in which all verbs occupy a position at the sec-

ond pole from constructions in which (some of) the verbs appear in the Nachfeld (e.g. in the third construction) or in the Mittelfeld (e.g in constructions with predicative infinitives or fixed expressions).

Besides a discussion of verb cluster formation in Dutch, this chapter considered cluster formation in German as well. While the two languages show several similarities regarding verb clusters, they also differ on a number of points. Similar to Dutch, the German IPP verbs are a subset of the clustering verbs, but in comparison to Dutch the amount of IPP verbs is much smaller. Moreover, the IPP test is not a necessary test for clustering in German, as some verbs that never occur as IPP may also occur as a clustering verb. In addition, the canonical word order in German verb clusters is the mirror image of the canonical word order in Dutch clusters.

This chapter presented a descriptive overview of the phenomena related to cluster formation. Chapter 2 and 3 will present how those phenomena are canonically analysed in respectively transformational grammar and monostratal grammar.

While several authors provide a list of verbs that can be used in verb clusters and/or the third construction, it turns out that they do not entirely agree on the clustering abilities of all verbs mentioned in the literature. The corpus study in chapter 5 will investigate whether the lists provided in this chapter are empirically valid. Moreover, the word order variation within the cluster will be investigated in order to verify whether the generalizations and constraints regarding word order variation proposed by a.o. Broekhuis & Corver (2015) accurately describe Dutch verb clusters.

CHAPTER 2

Status quæstionis: Transformational grammar

Transformational grammar is a branch of generative grammar that starts from the assumption that each sentence in a language has multiple levels of representation. Syntactic structures are derived by means of syntactic operations between the levels of representation.

This chapter presents the main transformational treatments dealing with Dutch verb clusters. Even though the theoretical analysis of verb clusters in this thesis will not be formulated in terms of transformational grammar, an overview of the transformational accounts is useful for two reasons. On the one hand, the first theoretical analyses of the topic were formulated in a transformational framework. On the other hand, the terminology used to refer to the phenomena related to cluster formation within descriptive literature and other theoretical frameworks refers to concepts that were developed within this framework. The purpose of this chapter is thus to present some influential analyses of verb clusters within transformational grammar.

The literature on the analysis of verb clusters within transformational grammar is vast. For an overview, see Wurmbrand (2001) and Wurmbrand (in press). Pioneering work on Dutch verb clusters is conducted by Evers (1975). His work is modified and extended by amongst others den Besten & Rutten (1989), and Haegeman & van Riemsdijk (1986). A summary of their main arguments will be given, as well as an alternative proposal of Zwart (2011) developed within the Minimalist Program. Section 2.1 discusses the derivation of verb clusters and extraposition. Section 2.2 discusses the third construction and verb clusters containing non-verbal elements.

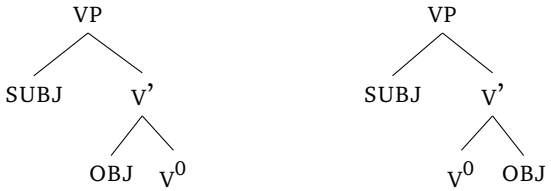
The linguistic examples in this chapter are generally taken from the cited authors, although some examples are slightly adapted in order to make the comparison between the different analyses more transparant.[1]

2.1 Extraposition versus V-Raising

This section presents the derivation of Dutch verb clusters, as opposed to constructions with extraposition. What the different analyses have in common, is the fact that the verbs and their arguments form a syntactic and semantic unit in the base structure. In order to derive verb clusters, the transformations typically involve the movement of verbs and/or their arguments.

A major distinction can be drawn between analyses that assume a head-final structure for Dutch, and those starting from a head-initial base structure.

(2.1) (a) Head-final structure (b) Head-initial structure



In the head-final approach the verb is base-generated to the right of its complements (2.1a), whereas in the head-initial approach it occurs to the left of its complements in the base structure (2.1b). Under the head-final hypothesis verb clusters are typically derived by means of rightward movement of the embedded verb or verb phrase, while in a head-initial analysis it involves the leftward movement of the complements (Wurmbrand 2005: 232–233).

In section 2.1.1 the head-final proposal of Evers (1975, 2004) and Haegeman & van Riemsdijk (1986) will be sketched, while in section 2.1.2 the head-initial alternative proposed by Zwart (1994, 2011) will be presented.

2.1.1 Head-final approaches

Evers (1975, 2004) starts from the assumption that Dutch is head-final (SOV), which

[1] For some examples this includes the addition of labeled bracketing. Another adaptation concerns the omission of the silent arguments PRO (for controlled subjects) and *e* (for raised subjects), as they are not relevant in this discussion.

means that in the base-generated structure all objects are located to the left of the matrix verb. In verb-final clauses this is the canonical position for nominal (2.2a) and adjectival complements (2.2b).

(2.2) a. ... dat hij meestal de bus neemt.
... that he usually the bus takes
'... that he usually takes the bus.'

b. Het lijkt alsof ze gelukkig is.
It seems as-if she happy is
'It seems like she is happy.'

By contrast, clausal complements are moved towards the end of the sentence, as shown in examples (2.3) and (2.4). The a-examples show the base-generated structure, whereas the b-examples show the derived structure.[2] This rightward movement is also known as *extraposition* (Evers 2004: 95).

(2.3) a. dat ik [dat hij me kent] denk
that I that he me knows think

b. dat ik t_1 denk [dat hij me kent]$_1$
that I think that he me knows
'that I think that he knows me.'

(2.4) a. dat hij [een fout gemaakt te hebben] merkte
that he a mistake made to have noticed

b. dat hij t_1 merkte [een fout gemaakt te hebben]$_1$
that he noticed a mistake made to have
'that he noticed he had made a mistake.'

Extraposition is not only applicable to finite complements, like *dat hij me kent* 'that he knows me' in (2.3), but also to non-finite complements, such as *een fout gemaakt te hebben* 'to have made a mistake' in (2.4).

The examples above show that extraposition does not lead to verb clusters nor to cross-serial dependencies, since the arguments are still adjacent to their verbs after the entire complement has moved. For the formation of verb clusters, Evers makes use of a different syntactic rule: *Verb Raising* (or *V-Raising*), illustrated in (2.5).

(2.5) a. ... dat [Piet twintig boterhammen eten] kan.
... that Piet twenty sandwiches eat can

[2]The trace *t* encodes the base-position of the moved element.

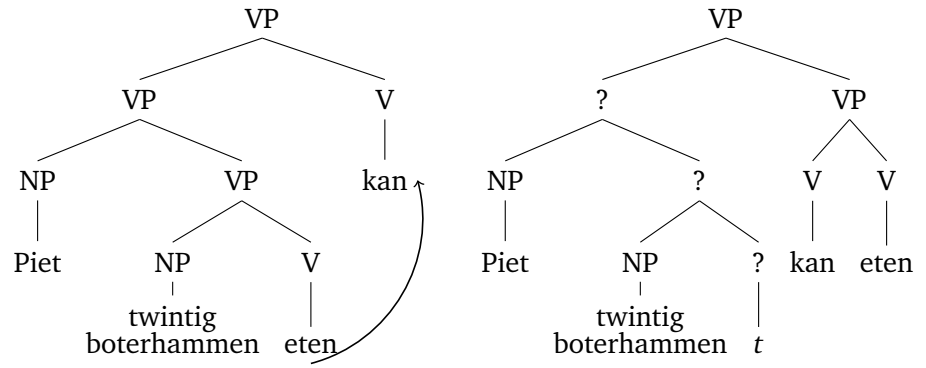
b. ... dat [Piet twintig boterhammen t_1] [kan eten$_1$]
... that Piet twenty sandwiches can eat'
'... that Piet is able to eat twenty sandwiches.'

c. * dat kan [Piet twintig boterhammen eten]
that can Piet twenty sandwiches eat

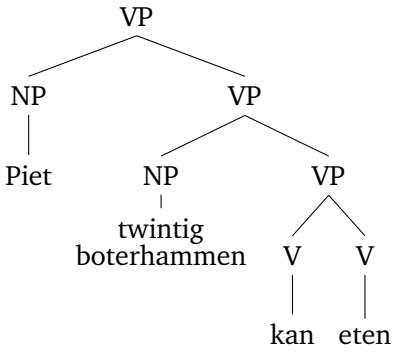
In those constructions, the non-finite verb is moved out of the complement and right-adjoined to the verb in the matrix clause, giving rise to a verb cluster, cf. *eten* 'eat' in (2.5a–b) (Evers 2004: 100). Note also that, in this example, extraposition leads to an ungrammatical sentence (2.5c), which means that *kunnen* obligatorily triggers verb raising.

After the verb has moved out of the non-finite complement, its arguments remain ungoverned (2.6a). As a solution, Evers prunes the nodes on the projection line of the verb's trace and moves the arguments to a position in the matrix clause (2.6b). This reattachment mechanism is called *argument relicensing* (Evers 2004: 98).

(2.6) a.



b.



Evers points out that the verbs triggering verb raising can be distinguished from verbs that trigger extraposition by means of the IPP effect: In the perfect tense the V-raisers always appear as an infinitive instead of a past participle. In example (2.7) the infinitive *kunnen* 'can' is used instead of the (expected) past participle *gekund*.

(2.7) a. dat [[hij de vogels met een gebaar de lucht in jagen] kunnen] heeft
that he the birds with a gesture the air in chase can has

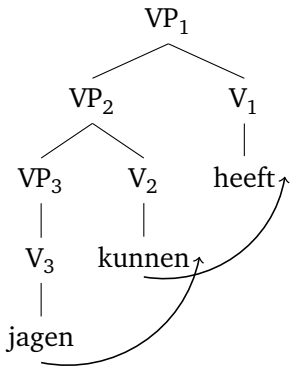
b. dat [[hij de vogels met een gebaar de lucht in t_1] [kunnen jagen$_1$]] heeft
that he the birds with a gesture the air in can chase has

c. dat [[hij de vogels met een gebaar de lucht in t_1]t_2] [heeft [kunnen jagen$_1$]$_2$]
that he the birds with a gesture the air in has can chase

'He was able to chase the birds into the air with a gesture.'

(2.7) shows that the application of verb raising is cyclic, as a verb that triggers verb raising can be raised itself. The verb *kunnen* triggers the raising of *jagen* 'chase', and the auxiliary *heeft* triggers the raising of the cluster *kunnen jagen*. A visual representation of the movement is presented in (2.8).

(2.8)



Besides the occurrence of the IPP effect, Evers states that verb raising and extraposition can be distinguished by some other criteria (Evers 2004: 102):

EXTRAPOSITION [– IPP]	V-RAISING [+ IPP]
- infinitival prefix *te*	- no infinitival prefix *te*
- initial complementizer *om* allowed	- *om* is not allowed
- a controlled subject	- never a controlled subject
- never subject raising	- always subject raising

Table 2.1: Extraposition versus V-raising (Evers 2004: 102)

There are a number of exceptions to these general criteria. For example, the verbs *proberen* 'try' and *schijnen* 'appear' take infinitival complements with the prefix *te*,

while allowing verb raising. According to Rutten (1991) the set of verbs that select a *te*-infinitive and obligatorily appear as IPP consists of fifteen verbs,[3] whereas the set of optional IPP verbs that select a *te*-infinitive consists of ten verbs.[4] Some examples are given in (2.9–2.10)

(2.9) . . . dat hij dat boek heeft weten/*geweten te vinden!
. . . that he that book has manage.IPP/*manage.PSP to find
'. . . that he managed to find that book!'

(2.10) a. . . . dat hij het boek heeft proberen te lezen.
. . . that he the book has try.IPP to read

b. . . . dat hij heeft geprobeerd het boek te lezen.
. . . that he has try.PSP the book to read
'. . . that he tried to read the book.'

The criteria for raised and controlled subjects turn out to be too strict as well. The subject control verb *proberen* 'try' may occur both as a V-raiser (2.10a) or an extraposition inducer (2.10b) if it selects another infinitive.

Evers' original verb raising approach gained a lot of attention and the analysis was adopted by several other linguists working on Germanic verb clusters, cf. amongst others den Besten & Edmondson (1983) and Rutten (1991).

Haegeman & van Riemsdijk (1986) propose an alternative way of deriving verb clusters from head-final structures. According to their analysis, verb cluster formation is a two-step process involving *reanalysis* and *inversion*, as illustrated in (2.11) (Haegeman & van Riemsdijk 1986: 422).

(2.11) a. dat hij [[het probleem te begrijpen] probeert]
that he the problem to understand tries

b. dat hij [het probleem [te begrijpen probeert]] (reanalysis)
that he the problem to understand tries

c. dat hij [het probleem [probeert te begrijpen]] (inversion)
that he the problem tries to understand
'that he tries to understand the problem.'

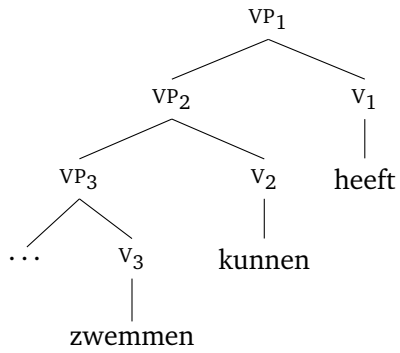
[3] *Blijken* 'appear', *dienen* 'be obliged to', *hebben* 'have to', *hoeven* 'need to', *horen* 'ought to', *komen* 'come', *lijken* 'seem', *liggen* 'lie', *lopen* 'walk', *plegen* 'be used to', *schijnen* 'appear', *staan* 'stand', *weten* 'know', *zien* 'manage', and *zitten* 'sit' (Rutten 1991: 29).

[4] *Beginnen* 'begin', *dreigen* 'threaten', *durven* 'dare', *helpen* 'help', *leren* 'learn; teach', *menen* 'mean, think', *proberen* 'try', *trachten* 'try', *wagen* 'dare to', and *weigeren* 'refuse' (Rutten 1991: 79).

In (2.11b) the base structure containing an embedded infinitival complement is reanalysed to a structure in which the verbs form a constituent.[5] In the next step, the positions of the two verbs - which are now sister nodes - are flipped. According to the authors the reanalysis+inversion approach has the advantage that it can deal with a larger range of constructions compared to the constructions covered by Evers' verb raising analysis, such as constructions in which non-verbal elements appear in the verb cluster (cf. section 2.2).

The proposal presented in Haegeman & van Riemsdijk (1986) was very influential as well. Wurmbrand (2004: 64–67), for instance, shows how inversion can be employed to account for the word order variation within West-Germanic three-verb clusters. Under the head-final hypothesis, the basic 3-2-1 order (e.g. *zwemmen kunnen heeft* 'swim can has') can be represented as the tree in (2.12).

(2.12)



The 1-3-2 order can be derived from the basic 3-2-1 order by inverting the hierarchically highest verb with its sister node. The 2-3-1 and 1-2-3 orders can be derived by inverting sister nodes as well. In order to derive the 2-1-3 and the 3-1-2 orders, however, inversion alone is not sufficient. With respect to the 2-1-3 order this is an advantage of the theory, since that word order is ungrammatical in all West-Germanic dialects. In order to account for the 3-1-2 word order, Wurmbrand employs a combination of inversion and leftward movement.[6] This type of reordering mechanism somewhat resembles the linearization constraints for verb clusters formulated in monostratal frameworks, cf. chapter 3.

[5]Haegeman & van Riemsdijk (1986) mention that the reanalysis is similar to the syntactic reanalysis of *John [talked [to Mary]]* to *John [[talked to] Mary]* in order to explain constructions like *Mary was talked to* (Haegeman & van Riemsdijk 1986: 420–421).

[6]In Dutch IPP constructions the 3-1-2 order is also ungrammatical, but the word order is accepted in modal-modal-verb and modal-auxiliary-verb combinations, e.g. *gemaakt moet hebben* 'made must have', cf. Table 1.4, based on Barbiers (2005).

2.1.2 A head-initial proposal

In contrast to treatments of cluster formation that start from the assumption that Dutch has an SOV structure, Zwart (1994, 2011) argues in his minimalist treatment that Dutch is head-initial, i.e. that it has an SVO structure. Similar to the approaches described in section 2.1.1, the arguments are adjacent to their verbs in the base structure (2.13a).

In order to derive SOV order in verb-final sentences only the non-verbal arguments are moved, which implies that the position of the verbs remains unchanged in clusters with a canonical word order. An example of such *object shift* is (2.13b) (Zwart 2011: 313).

(2.13) a. dat [Tasman probeert [te ontdekken het Zuidland]]
that Tasman tries to discover the South Land

b. dat [Tasman [het Zuidland]$_1$ probeert [te ontdekken t$_1$]]
that Tasman the South Land tries to discover
'that Tasman tries to discover the South Land.'

Zwart argues that no real *cluster formation* takes place in constructions like (2.13b), since the verb cluster is the residue out of which all non-verbal material is extracted (Zwart 2011: 313). In order to account for word order variation in the cluster, the verbs can move more to the left as well, for instance to derive verb-auxiliary combinations like *gelezen*$_2$ *heeft*$_1$ 'has read' (Zwart 1996).

An advantage of Zwart's approach is that only one mechanism is used to analyse constructions with and without verb clusters, i.e. movement to the left. In order to derive constructions with an extraposed infinitival VP, the complement is also moved to the left, but it stays within the embedded phrase. An example is (2.14), which is also derived from the structure in (2.13a).

(2.14) dat [Tasman probeert [[het Zuidland]$_1$ te ontdekken t$_1$]]
that Tasman tries the South Land to discover
'that Tasman tries to discover the South Land.'

Note that the derived structure in (2.14) looks like extraposition, but while extraposition (in its technical sense) involves rightward movement, this is not the case in Zwart's approach.

Compared to the transformational analyses proposed by Evers, Zwart only makes use of *object shift*, while an analysis in the line of Evers requires a mechanism to derive

extraposition and another one to derive verb raising. Moreover, his analysis does not have the relicensing problem, as the verbs are not moved out of their projections. Zwart furthermore argues that his treatment of verb clusters can easily account for a larger variation of constructions with verb clusters with less rules, such as clusters containing non-verbal elements (cf. section 2.2).

2.2 Third Construction and VP-Raising

This section presents transformational accounts for the derivation of the third construction and clusters containing non-verbal elements. Similar to the discussion on extraposition and verb raising in section 2.1, a major distinction can be drawn between head-final and head-initial analyses. A discussion of the former is presented in section 2.2.1, while section 2.2.2 presents a head-initial account.

2.2.1 Head-final approaches

According to den Besten et al. (1988) and den Besten & Rutten (1989), the dichotomy between extraposition and verb raising as proposed in Evers (1975) is not sufficient to cover all the Dutch data containing sentential embedding. They show that the appearance of a verbal sequence is not a sufficient condition for verb raising, since there exists a set of verbs that allow the formation of a verbal sequence without (necessarily) triggering IPP, cf. *adviseren* 'advise' in example (2.15). They refer to this kind of construction with the term *third construction* (den Besten et al. 1988: 17).

(2.15) a. ... dat hij t_1 heeft geadviseerd [het boek te lezen]$_1$
... that he has advised.PSP the book to read

b. ... dat hij [het boek]$_2$ t_1 heeft geadviseerd [t_2 te lezen]$_1$
... that he the book has advised.PSP to read
'... that he has advised to read the book.'

Den Besten et al. (1988) claim that the third construction is a combination of extraposition (2.15a) and the raising of (one or more) non-verbal arguments (2.15b). The movement of those arguments is also known as *scrambling*. As a result of the derivations a sequence of verbs appears, but the IPP effect does not show up. Similar to verb raising constructions, (at least some of) the complements are separated from their verbs, which leads to crossing dependencies.

Note that optional IPP verbs such as *proberen* 'try' are ambiguous between the verb raising construction and the third construction if they occur in the finite form.

(2.16) a. ... dat hij het boek probeert te lezen
... that he the book tries to read
... that he tries to read the book.'

b. ... dat hij [het boek t_1] probeert [te lezen]$_1$

c. ... dat hij [het boek]$_2$ t_1 probeert [t_2 te lezen]$_1$

The construction in (2.16a) might result from the raising of *te lezen* 'to read' (2.16b), but an alternative analysis is the extraposition of the verbal complement and the scrambling of the NP *het boek* 'the book' to the Mittelfeld (2.16c).

In his later work, Evers takes into account the third construction as well. He argues against the use of the scrambling rule, since there are several unsolved problems regarding scrambling, such as the trigger, the target, and the scope of the rule (Evers 2004: 96). As an alternative he extends his treatment of extraposition (see section 2.1) to the third construction, which he treats as a *partial extraposition*. In order to get the third construction with a verbal sequence starting from the structure in (2.17a),[7] only the head of the embedded complement is moved while all the remaining arguments are relicensed (2.17b). In the constructions (2.17c–e) not only the head is extraposed, but some of the non-verbal arguments as well. This is possible via the *Pied Piping* mechanism: The verb's projection line is partially moved along with the verbal head, which explains the name *partial extraposition*.[8] In the case of full extraposition (2.17f), the complete projection line is pied-piped, so none of the non-verbal arguments remain in situ (Evers 2004: 95–96).

(2.17) a. als hij [de vogels met een gebaar de lucht in te jagen] [had geprobeerd]
if he the birds with a gesture the air in to chase had tried.PSP
'if he had tried to chase the birds into the air with a gesture.'

b. als hij [de vogels met een gebaar de lucht in t_1] [had geprobeerd [te jagen]$_1$]

c. als hij [de vogels met een gebaar de lucht t_1] [had geprobeerd [in te jagen]$_1$]

d. als hij [de vogels met een gebaar t_1] [had geprobeerd [de lucht in te jagen]$_1$]

[7]Note that the structure in (2.17a) is already a derived structure, since the finite verb occurs to the left of the past participle.

[8]The notion of *Pied Piping* is often discussed in the context of *wh*-constructions, in which the movement of the *wh*-element triggers the movement of other non-*wh*-elements (Evers 2004: 93-94). Pied Piping of *wh*-elements can be obligatory, as in *Which car did he buy __?* versus * *Which did he buy __ car?*, or optional, as in *To whom did you give the book __?* versus *Whom did you give the book to __?*.

e. als hij [de vogels t_1] [had geprobeerd [met een gebaar de lucht in te jagen]$_1$]

f. als hij [t_1] [had geprobeerd [de vogels met een gebaar de lucht in te jagen]$_1$]

The difference between verb raising and the third construction lies in the type of verbal element that is moved: In the case of verb raising, the head of a bare verb phrase is moved, while in the case of the third construction, a clause (IP) is moved.

Evers claims that deriving partial extraposition by Pied Piping has the advantage that the linear ordering of the elements that are moved out is kept intact, which is not straightforwardly the case in the scrambling analysis proposed in den Besten et al. (1988).

While the third construction shows that the occurrence of an uninterrupted verbal sequence is not a sufficient condition for verb raising, there is a phenomenon that indicates it is not a necessary condition either. Haegeman & van Riemsdijk (1986) describe such cases of *Verb Projection raising* (or *VP-raising*) as constructions with IPP, but without a verb cluster. Or rather, the verb cluster is interrupted by non-verbal elements. An example of VP-raising in West Flemish is given in (2.18), derived from the base-generated structure in (2.18a) (Haegeman 1988: 672-674).

(2.18) a. da Jan [[Valère [[geen vlees] eten]] liet]
that Jan Valère no meat eat let

b. da Jan [Valère [[[geen vlees] eten] liet]] (reanalysis)
that Jan Valère no meat eat let

c. da Jan [Valère [liet [[geen vlees] eten]]] (inversion)
that Jan Valère let no meat eat
'that Jan allowed that Valère did not eat any meat.'

VP-raising is considered to be a variation on verb raising, also involving reanalysis and inversion.[9] The verb *liet* 'let' is flipped over the phrase *geen vlees eten* 'eat no meat', resulting in an interrupted verb cluster.

If a VP-raising construction contains an auxiliary of the perfect, the IPP effect shows up, but the verb cluster is interrupted (2.19a). The equivalent in Standard Dutch (2.19b) is not grammatical. However, VP-raising occurs in Standard Dutch as well, especially in the variant that is spoken in Belgium (cf. section 1.6). For example if the non-verbal elements are separable verb particles, such as *bezig* 'busy' and *na*

[9] In the verb raising variant the constructions would be reanalysed differently: The verbs would form a unit before combining with the NP complements, in a similar fashion as in (2.11).

'after' in (2.20). Evers calls such non-verbal elements *cluster creepers* (Evers 2003: 43).

(2.19) a. WV: da Jan hee willen *en hus* kopen
that Jan has want a house buy

b. NL: * dat Jan heeft willen *een huis* kopen
that Jan has want a house buy
'that Jan has wanted to buy a house.'

(2.20) a. ... dat ik me daar niet mee heb willen *bezig* houden
... that I me there not with have want busy keep
'... that I did not want to bother with that.'

b. ... dat ik daar niet meer over zal moeten *na* denken
... that I there not again about will have-to after think
'... that I will not have to think about that again.'

In his revised theory, Evers treats the derivation of V-raising and VP-raising in a similar fashion to the construction of (partial) extraposition (Evers 2004: 99).

(2.21) a. ... omdat hij jullie [de vogels met een gebaar de lucht in jagen] liet
... because he you the birds with a gesture the air in chase let
'... because he let you chase the birds with a gesture into the air'

b. ... omdat hij jullie [de vogels met een gebaar de lucht in t_1] [*liet* [jagen]$_1$]

c. ... omdat hij jullie [de vogels met een gebaar de lucht t_1] [*liet* [in jagen]$_1$]

d. ... omdat hij jullie [de vogels met een gebaar t_1] [*liet* [de lucht in jagen]$_1$]

e. ... omdat hij jullie [de vogels t_1] [*liet* [met een gebaar de lucht in jagen]$_1$]

f. ... omdat hij jullie [t_1] [*liet* [de vogels met een gebaar de lucht in jagen]$_1$]

g. *... omdat hij t_1 [*liet* [jullie de vogels met een gebaar de lucht in jagen]$_1$]

Starting from the structure in (2.21a), only the verbal head is right-adjoined in order to derive a verb raising construction (2.21b). The non-verbal arguments all remain in situ and are *relicensed* (cf. section 2.1). For (partial) VP-raising, parts of the embedded complement are pied-piped (2.21c–e). Example (2.21f) shows an example of *full* VP-raising, in which the entire embedded complement has moved towards the end of the clause. (2.21g) is ruled out as the external argument of the embedded complement (in this case *jullie* 'you') cannot be pied-piped along with the embedded complement. Evers labels the (partial and full) VP-raising constructions in (2.21) as 'southern Dutch'. He argues that in northern Dutch Pied Piping of non-verbal complements is blocked, as that variant requires the adjacency of the matrix verb and the head of the embedded complement.

2.2.2 A head-initial proposal

In his head-initial analysis of Duch, Zwart (2011) derives the third construction and VP-raising in a similar way to the derivation of verb raising and extraposition, i.e. by moving the non-verbal elements to the left. Starting from the argument structure in (2.22a), constructions with an interrupted cluster can be derived by moving the non-verbal complements only part of the way (2.22b), instead of all the way up to the Mittelfeld (Zwart 2011: 314–315).

(2.22) a. dat [Tasman wou [proberen [te ontdekken het Zuidland]]]
that Tasman wanted try to discover the South Land

b. dat [Tasman wou [[het Zuidland]$_1$ proberen [te ontdekken t$_1$]]]
that Tasman wanted the South Land try to discover
'that Tasman wanted to try to discover the South Land.'

Instances of the third construction can be derived by the same mechanism, so Zwart does not assume a combination of extraposition and scrambling in order to analyse such constructions. What needs to be explained, though, is the difference between constructions with verb clusters, showing the IPP effect in the perfect tense (2.23a), and instances of the third construction, in which the IPP effect does not show up (2.23b).

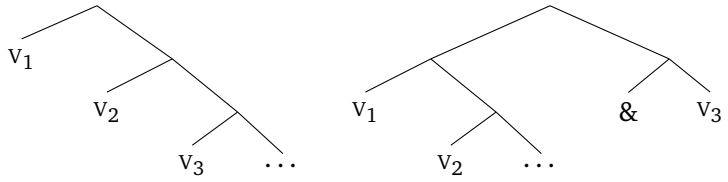
(2.23) a. dat Tasman het Zuidland heeft proberen te ontdekken
that Tasman the South Land has try.IPP to discover

b. dat Tasman het Zuidland heeft geprobeerd te ontdekken
that Tasman the South Land has tried.PSP to discover
'that Tasman has tried to discover the South Land.'

Zwart (2007) argues that the difference between verb raising and the third construction can be expressed as a difference between subordination and coordination. In a verb raising construction the verbs form a construction in which the verb of the embedded complement is dominated by another verb in the cluster, whereas in the third construction the relation between the verb of the embedded complement and the other verbs in the constructions is one of coordination, rather than subordination, i.e. the extraposed infinitive is juxtaposed to the verb cluster (Zwart 2007: 84).

The difference between both constructions is shown in (2.24). (2.24a) presents the verb raising construction, whereas a representation along the lines of (2.24b) presents the structure for the third construction.

(2.24) (a) Verb raising (b) Third construction



Ter Beek (2008) states that the difference between verb raising and the third construction lies in the syntactic type of embedded complement. In the case of extraposition the complement clause is a CP, and in the case of verb raising it is a VP. In the third construction it is not a CP, as complementizers are not allowed, but it is larger than a VP (e.g. a TP if the embedded complement has its own temporal interpretation) (Ter Beek 2008: 171–173).

2.3 Conclusion

This chapter has presented the most influential transformational treatments of verb cluster formation in Dutch. A major distinction was made between analyses based on the assumption that Dutch is head-final, such as the ones proposed by Evers (1975, 2004), den Besten & Rutten (1989), and Haegeman & van Riemsdijk (1986), and the analyses in the line of Zwart (2011), which start from the assumption that Dutch is verb-initial.

The head-final analyses involve rightward movement of the verbs and leftward movement of the arguments in order to derive verb clusters, whereas in the head-initial treatment only leftward movement is employed. Despite the formal similarity between verb raising and the third construction, the former make use of different mechanisms in order to derive V-raising/VP-raising and extraposition/third construction. In Zwart's proposal, the different constructions are derived in a more uniform way.

Table 2.2 provides an overview of the relation between the occurrence of a verbal sequence and IPP, as well as the terminology used for the constructions that were discussed in this section.[10]

[10]The terminology varies among different authors, such as the *third construction* in den Besten & Rutten (1989) versus *partial extraposition* in Evers (2004).

	+ IPP	− IPP
+ ADJACENT	(a) ...dat Cecilia het nijlpaard frietjes heeft proberen te voeren. *V-raising*	(b) ...dat Cecilia het nijlpaard frietjes heeft geprobeerd te voeren. *Third construction/Partial extraposition*
− ADJACENT	(c) ...dat Cecilia het nijlpaard heeft proberen frietjes te voeren. *(Partial) VP-raising* (e) ...dat Cecilia heeft proberen het nijlpaard frietjes te voeren. *(Full) VP-raising*	(d) ...dat Cecilia het nijlpaard heeft geprobeerd frietjes te voeren. *Third construction/Partial extraposition* (f) ...dat Cecilia heeft geprobeerd het nijlpaard frietjes te voeren. *(Full) extraposition*

Table 2.2: Relation IPP – adjacency of the verbs

The constructions with IPP, i.e. the constructions with V/VP-raising, are typically considered to be *genuine* verb clusters, whereas the constructions without IPP are not.

CHAPTER 3

Status quæstionis: Monostratal grammar

In contrast to transformational grammar, monostratal approaches such as Categorial Grammar (CG), Generalized Phrase Structure Grammar (GPSG), and Head-driven Phrase Structure Grammar (HPSG) take into account only one layer of analysis, which corresponds to the surface structure in multistratal frameworks. In those frameworks, the central question with regard to verb cluster formation is how to link the arguments to the correct verb.

This chapter starts off with a brief sketch of the CG and GPSG analyses of verb clusters (section 3.1), since they are important sources of inspiration for the later HPSG analysis. Next, the HPSG framework will be introduced (section 3.2), as it is the framework that will be used throughout the rest of this thesis. Section 3.3 presents the HPSG analysis of verb clusters on which most of the current analyses within the framework are based, i.e. the *argument inheritance* mechanism proposed by Hinrichs & Nakazawa (1994). Sections 3.4 and 3.5 present how the argument inheritance analysis was further extended and refined in later work on German and Dutch verb clusters. The examples provided in this chapter are based on the literature, or constructed analogous to the examples provided by the authors under discussion.

3.1 Cluster formation in CG and GPSG

The two analyses of cluster formation presented in this section can be considered as the predecessors of the later HPSG proposals. First, the CG analysis proposed in

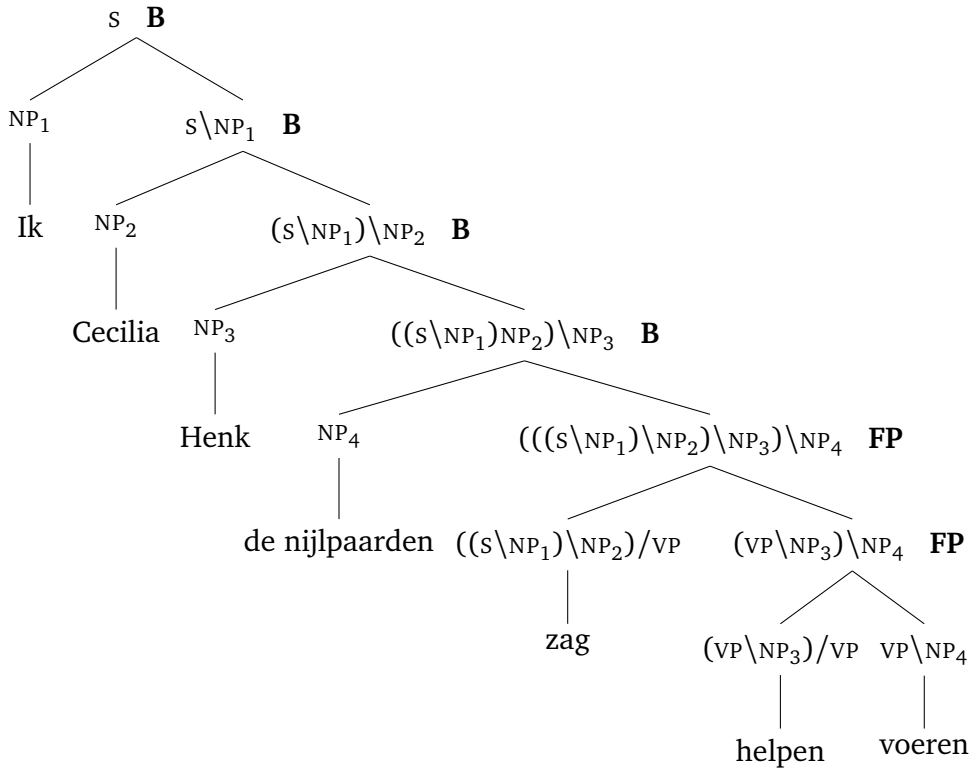
Steedman (1985) will be sketched, followed by the GPSG account of German verb clusters developed in Johnson (1986).

3.1.1 Categorial Grammar (CG)

The analysis of Dutch verb clusters, as proposed in Steedman (1985, 2000), can be illustrated as a binary branching tree structure. The verbs form a cluster to which the non-verbal arguments are added one by one. For instance, the example in (3.1), can be represented as (3.2), following (Steedman 2000: 141).[1]

(3.1) ...omdat ik Cecilia Henk de nijlpaarden zag helpen voeren.
...because I Cecilia Henk the hippos saw help feed
'...because I saw Cecilia help Henk feed the hippos.'

(3.2)



In CG *functional application* refers to the mechanism in which a category of type X/Y combines with category Y in order to get a mother node of type X. Functional

[1]The representation of the structure in Steedman (2000) is slightly different, but it is equivalent to the representation in (3.2). Steedman's notation S_{+SUB} (for subordinate clause) is shortened to S and VP_{-SUB} to VP in order to make the combinatorics more transparent.

application includes forward combination (F), and backward combination (B):

Forward combination: X/Y Y ⇒ X
Backward combination: Y X\Y ⇒ X

In the construction in (3.2) backward combination is used in order to attach the NPs in order to get a saturated S.

In order to deal with Dutch verb clusters, an additional mechanism is needed, i.e. *functional composition*. Functional composition extends the regular combinatory rules by combining categories of type X/Y with categories of type Y/Z to get a mother node X/Z. In that way, the list of categories that Y needs to get saturated can be propagated to higher nodes. Functional composition includes forward partial combination (FP) and backward partial combination (BP):

Forward partial combination: X/Y Y\Z ⇒ X\Z
Backward partial combination: Y\Z X\Y ⇒ X\Z

Forward partial combination is used in (3.2) for the construction of the verb cluster.

3.1.2 Generalized Phrase Structure Grammar (GPSG)

Within the GPSG framework, Johnson (1986) claims that there are two possible analyses of the German VP constituent structure. In the first analysis the main verb *sehen* 'see' and its non-verbal arguments combine before the auxiliary *haben* 'have' is added (3.3a). The main verb and its arguments thus form a syntactic and a semantic unit. In the second analysis, the main verb combines with the auxiliary before adding the objects (3.3b). This is more in line with the CG analysis described in the previous section.

(3.3) a. Fritz sagte, dass er [[Helmut gesehen] hatte].
Fritz said that he Helmut seen had

b. Fritz sagte, dass er [Helmut [gesehen hatte]].
Fritz said that he Helmut seen had
'Fritz said that he had seen Helmut.'

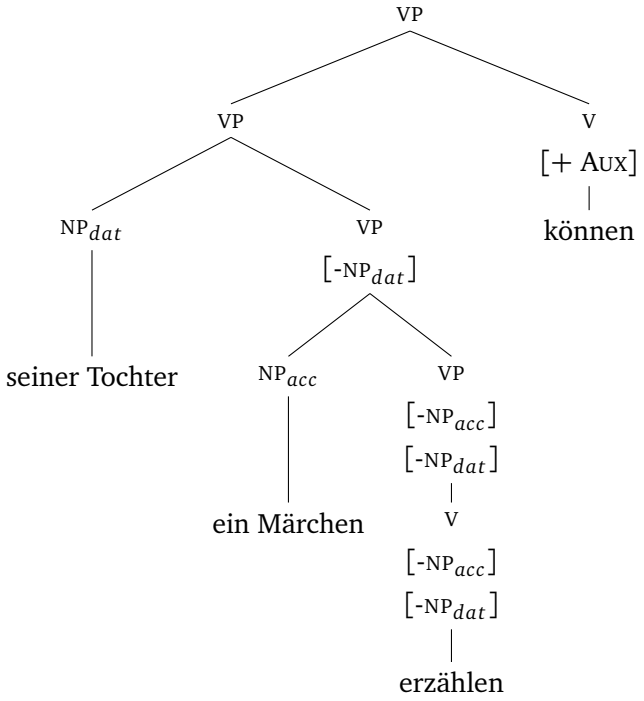
The second analysis is based on the observation that verb clusters behave like other constituents. In German V-initial clauses only one constituent can be placed in front of the finite verb. Making use of the fronting test, Johnson concludes that *partial VPs* are constituents, e.g. *erzählen können* in (3.4b) (Johnson 1986: 873–874).

(3.4) a. Er hat seiner Tochter ein Märchen erzählen können.
he has his daughter.DAT a tale.ACC tell can

b. *Erzählen können* hat er seiner Tochter ein Märchen.
tell can has he his daughter.DAT a tale.ACC
'He has been able to tell his daughter a tale.'

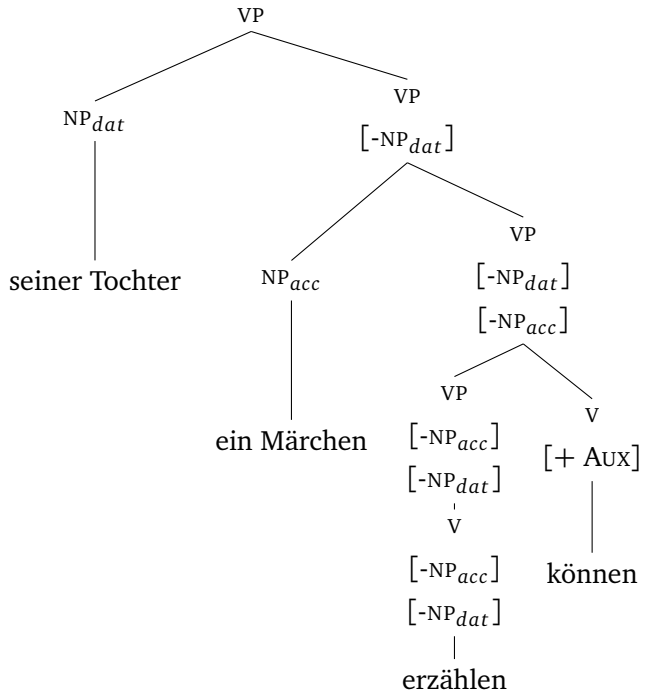
In the representation of tree structures, Johnson uses *minus category features*, such as [-NP$_{acc}$], to indicate the subcategorization pattern of verbs. Those features are similar to the *slash categories* in CG: They indicate which complements a verb needs in order to form a saturated constituent (Johnson 1986: 874–875). The tree structure in (3.5a) shows the analysis for the VP in (3.4a), in which the verbs form a constituent with their objects. The structure in (3.5b) presents the alternative analysis in which a verb cluster is formed.[2] This analysis is necessary to be able to explain the construction in (3.4b). In both constructions the non-verbal arguments are attached to the cluster one by one, which is similar to the CG analysis (Johnson 1986: 877).

(3.5) a.



[2]Johnson makes use of the expansion rule VP → V to derive the non-branching VP structure for *erzählen* 'tell' (Johnson 1986: 875).

b.



Johnson's arguments to treat auxiliaries and main verbs as one constituent, as well as the stepwise combination of verbs with their arguments is adopted in the HPSG analysis of amongst others Hinrichs & Nakazawa (1994), which will be presented in section 3.3. Besides the *verb cluster analysis* in (3.5b), Johnson considers structures like (3.5a) in which verbs first combine with their non-verbal arguments as a valid alternative, but this analysis is not taken up in HPSG.

3.2 Introduction to HPSG

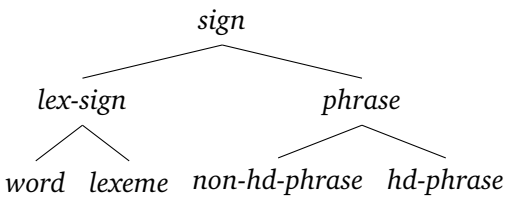
Before turning to the current HPSG analyses of the verb cluster in German and Dutch, this section presents a short introduction to the basic HPSG architecture, based on Pollard & Sag (1994), Ginzburg & Sag (2000), and Sag et al. (2003). This introductory section is included to make the linguistic analyses in the next chapters accessible to any linguist, regardless of his or her theoretical background. For a more comprehensive introduction to HPSG, the reader is referred to the works cited above.

3.2.1 Signs and types

HPSG and the transformational branch of generative grammar share a similar goal, i.e. to express the relation between form and meaning in language. The treatments differ in the way the relationship is expressed. Instead of capturing the relation by transformations of tree structures, only one structure is used: A *typed feature structure* (TFS).

Linguistic objects belong to a *type*. The most general type is that of the signs, which can be subdivided into subtypes along the hierarchy in (3.6). Subtypes inherit all properties of their supertype.

(3.6)



Signs can be lexical or phrasal. The latter will be discussed in section 3.2.2. The lexical signs (*lex-sign*) are subdivided into lexemes and words. *Words* are the basic components of syntax which are used to build phrases, whereas *lexemes* are used to abstract over word forms. For example, the lexeme *boom* 'tree' is an abstraction of the singular word form *boom* 'tree', the plural *bomen* 'trees', and the diminutive forms *boompje* 'small tree' and *boompjes* 'small trees'. The verbal lexeme *lopen* 'run' stands for the words *loop* 'run-1SG', *loopt* 'run-2/3SG', *lopen* 'run-PL', *liep* 'ran-SG', *liepen* 'ran-PL', *lopend* 'running-PRP' and *gelopen* 'run-PSP'.

Each type is associated with certain features. The relation between types and features, as well as the constraints associated with grammar rules, principles and lexical entries are expressed by *feature structure descriptions*, which are represented as attribute-value matrices (AVMs) (Sag et al. 2003: 193). Within an AVM, type labels are typeset in *italics*, whereas feature labels (or, equivalently, attribute labels) are typeset in SMALL CAPITALS. Feature values are also typeset in *italics*, as they are also types.

Objects of the type *sign* have the features PHON(OLOGY) and SYNTAX-SEMANTICS (SYNSEM or SS). PHON is a list of strings corresponding to the sign's utterance. SYNSEM contains the sign's syntactic and semantic information. This can be expressed in an AVM as follows:

(3.7)
$$\begin{bmatrix} \textit{sign} & \\ \textsc{phon} & \textit{list}\left(\textit{phoneme}\right) \\ \textsc{synsem} & \textit{synsem} \end{bmatrix}$$

The features PHON and SYNSEM are assigned to all linguistic signs, but there are also features that are only used for specific types of signs. For example, phrases contain the feature DAUGHTERS (DTRS), which indicates the main *building blocks* of the phrase. Lexical signs do not have such a feature, as they are syntactic atoms.

Furthermore, it is possible to embed one AVM into another. For example, the feature SYNSEM in (3.7) takes as its value a type *synsem*, which is associated with the features LOC for information on local phenomena and NONLOC for establishing nonlocal dependencies. Information that is relevant in a local context includes the syntactic CAT(EGORY) and the semantic CONT(ENT) of the sign. The feature structure in (3.7) can thus be extended to (3.8).

(3.8)
$$\begin{bmatrix} \textit{sign} & & \\ \textsc{phon} & \textit{list}\left(\textit{phoneme}\right) & \\ \textsc{synsem} & \begin{bmatrix} \textit{synsem} & & \\ \textsc{loc} & \begin{bmatrix} \textit{local} & \\ \textsc{cat} & \textit{category} \\ \textsc{cont} & \textit{semantic-object} \end{bmatrix} \\ \textsc{nonloc} & \textit{nonlocal} \end{bmatrix} \end{bmatrix}$$

Note that there is a difference between an AVM and a TFS. In an AVM, certain properties can be left underspecified, as it is a representation of a feature structure description. In a TFS, all values have to be spelled out.

The AVM in (3.8) can be further specified, for instance to the feature structure in (3.9). It represents the lexical entry *boom* 'tree', which contains information about the word's part of speech, its NUM(BER), and GEN(DER) in the HEAD feature. Note that the pipe (|) symbol can be used to abbreviate the notation: for example, the path LOC|CAT|HEAD is equivalent to the structure [LOCAL [CATEGORY [HEAD]]].

(3.9)
$$\begin{bmatrix} \textit{word} \\ \text{PHON} \quad \langle \text{boom} \rangle \\ \text{SYNSEM} \quad \begin{bmatrix} \textit{synsem} \\ \text{LOC} \mid \text{CAT} \mid \text{HEAD} \begin{bmatrix} \textit{noun} \\ \text{AGR} \begin{bmatrix} \text{NUM} & \textit{sg} \\ \text{GEN} & \textit{masc} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

3.2.2 Headed phrases

The HEAD value contains all features which the lexical head of a phrase shares with the complete phrase, such as the part-of-speech.

All phrases have daughters. Phrases of the type *h(eade)d-phrase* always have a HEAD DAUGHTER (HD-DTR) feature.[3] In headed phrases the HEAD features of the head daughter are identical to the head features of the mother. This is captured by the *Head Feature Principle* (Pollard & Sag 1994: 34), formulated as follows in Van Eynde (2006: 163):

(3.10) Head Feature Principle:

$$\textit{hd-phr} \quad \Rightarrow \quad \begin{bmatrix} \text{SYNSEM} \mid \text{CAT} \mid \text{HEAD} \boxed{1} \\ \text{HD-DTR} \mid \text{SYNSEM} \mid \text{CAT} \mid \text{HEAD} \boxed{1} \end{bmatrix}$$

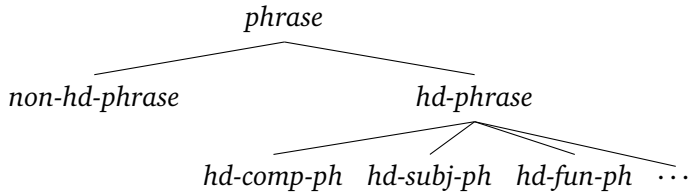
For example, in the noun phrase *de boom* 'the tree' the noun *boom* 'tree' shares its HEAD feature with the complete noun phrase

The boxed numbers in (3.10) are a means to indicate that the HEAD values of the mother node are identical to the HEAD values of the head daughter. This mechanism of *structure sharing* is characteristic for HPSG.

In order to model how a head combines with its syntactic arguments, phrasal constraints on headed phrases are used. The type *hd-phrase* in the hierarchy in (3.6) consists of several subtypes. The most important ones are spelled out in (3.11):

[3] Non-headed phrases, such as coordinate structures, lack this feature.

(3.11)



In this section, *head-complement phrases (hd-comp-ph)*, *head-subject phrases (hd-subj-ph)*, and *head-functor phrases (hd-fun-ph)* will be discussed.

Head-Complement Structures

The constraint on head-complement phrases models the combination of a head with its complements. The version of Müller (2002: 16) is adopted, as it allows for a stepwise combination with the complements.[4] It can be defined as follows:

(3.12)

$$\textit{hd-comp-ph} \Rightarrow \left[\begin{array}{l} \textsc{synsem}\,|\,\textsc{loc}\,|\,\textsc{cat}\,|\,\textsc{comps} \quad \boxed{A} \\ \textsc{hd-dtr}\,|\,\textsc{synsem}\,|\,\textsc{loc}\,|\,\textsc{cat}\,|\,\textsc{comps} \quad \boxed{A} \oplus \left\langle \boxed{1} \right\rangle \\ \textsc{nonhd-dtr}\,|\,\textsc{synsem} \quad \boxed{1}\ \textit{synsem} \end{array}\right]$$

The '⊕' stands for the *append* relation, which concatenates two lists. In a head-complement phrase the complement daughter is represented in the list of non-head daughters. Its SYNSEM value should be identical to the last element in the COMPS list of the head daughter (1). As the complement is realised, it can be cancelled off in the valence requirements of the mother (A). The remaining COMPS list may be empty, but it may also contain a list of *synsem* objects.[5] An illustration of the constraint for the verb phrase *hem een fles wijn geeft* 'gives him a bottle of wine' in (3.13), is shown in (3.14).[6]

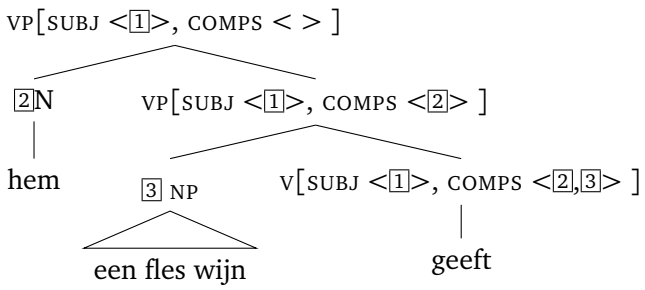
[4] (3.12) assumes that complements are added one at a time, from the most to the least oblique. Ginzburg & Sag (2000: 34) provide a slightly different version of the head-complement constraint, as they assume that the head combines with all its complements in one single step. Müller (2002) uses a NON-HEAD-DTRS feature. As the analysis presented in chapters 6 and 7 relies on binary branching structures, there is only one non-head daughter (NONHD-DTR). As a consequence, it can be referred to directly, rather than via a list.

[5] In English, the list in A has to be empty, since English is subject to the *Empty Comps Constraint* (Ginzburg & Sag 2000: 33).

[6] Note that instead of expanding the NP *hem* 'him' to a non-branching N-node, it is abbreviated to N. This abbreviation is applied to all NPs with an empty valence list, such as pronouns and proper names (Sag et al. 2003: 101).

(3.13) . . . dat ze hem een fles wijn geeft.
. . . that she him a bottle wine gives
'. . . that she gives him a bottle of wine.'

(3.14)



The non-verbal complements are canonically realised in the same order as they appear on the COMPS list of the verb. They are typically ordered from the least oblique to the most oblique complement, according to the *obliqueness hierarchy* proposed by Keenan & Comrie (1977).[7] *Geven* 'give' in (3.13) subcategorises for two NP complements, i.e. an indirect object (*hem* 'him') and a direct object (*een fles wijn* 'a bottle of wine'). The indirect object can also appear as a PP, in which case it canonically follows the direct object NP (e.g. *dat hij een fles wijn aan Jan geeft* 'thay he gives a bottle of wine to Jan'). This is encoded in a separate lexical entry for *geven*, which has <NP,PP> as a COMPS list.

Head-Subject Structures

The constraint on head-subject phrases models the combination of a head with its subject:

(3.15)

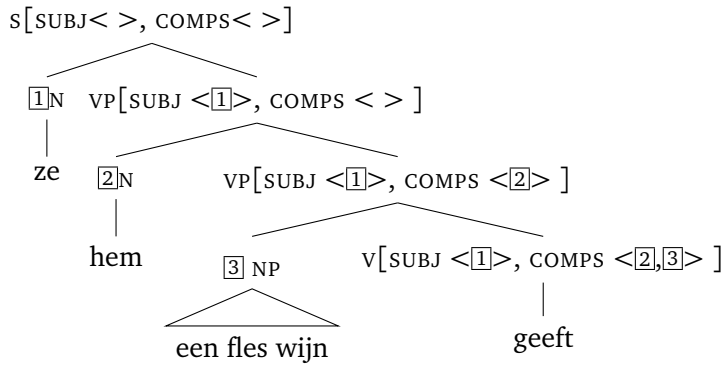
$$\textit{hd-subj-ph} \quad \Rightarrow \quad \begin{bmatrix} \text{SYNSEM} & \left[\text{LOC}\,|\,\text{CAT}\,|\,\text{SUBJ}\left\langle\ \right\rangle\right] \\ \text{HD-DTR} & \left[\text{SYNSEM}\,|\,\text{LOC}\,|\,\text{CAT}\,|\,\text{SUBJ}\left\langle \boxed{1} \right\rangle\right] \\ \text{NON-HD-DTR} & \left[\text{SYNSEM}\ \boxed{1}\ \textit{synsem}\right] \end{bmatrix}$$

[7] SUBJECT > DIRECT OBJECT > INDIRECT OBJECT > OBLIQUE CASE > GENITIVE > OBJECT OF COMPARISON. The elements in the hierarchy are ordered from least oblique (or most accessible) to most oblique (or least accessible). In older versions of HPSG only one SUBCAT list is used to express the subject and the complement requirements (cf. infra). In more recent treatments, the subject appears on a separate SUBJ list and the complements appear on the COMPS list.

In head-subject phrases, the SYNSEM value of the non-head daughter is identical to the value of the head daughter's SUBJ list (Ginzburg & Sag 2000: 34).

An example is the combination of the VP in (3.14) with its subject *ze* 'she' in (3.16).

(3.16)



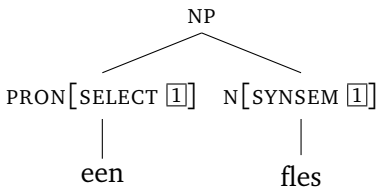
Head-Functor Structures

Head-functor phrases are structures in which the non-head (functor) daughter selects its head sister, as captured by the constraint in (3.17) (Van Eynde 1998: 130–131).

(3.17) $$\textit{hd-fun-ph} \Rightarrow \begin{bmatrix} \text{HD-DTR} \mid \text{SYNSEM} \boxed{1} \textit{ synsem} \\ \text{NONHD-DTR} \mid \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{HEAD} \mid \text{SELECT} \boxed{1} \end{bmatrix}$$

The SELECT feature is a head feature of words which replaces the formerly used SPEC(IFIED) and MOD(IFIED) features used by markers, determiners and adjuncts to select their head (Pollard & Sag 1994: 45–46,55). For example, in the NP *een fles* 'a bottle' the determiner *een* selects the noun *fles*, requiring it to be singular and count:

(3.18)



The head-functor phrase type generalizes over *head-adjunct*, *head-specifier*, and *head-marker* phrases, cf. Pollard & Sag (1994).

3.2.3 Argument selection versus argument realization

In order to model subcategorization in HPSG a distinction is made between argument selection and argument realization. *Argument selection* is a lexical property, which is

modeled on the ARG(UMENT)-ST(RUCTURE) list. The ARG-ST feature is assigned to all *lex-sign* types (i.e. words and lexemes), and its value is a list of SYNSEM values. For example, the fact that the verb *kijken* 'look' in (3.19) selects a noun phrase (*Bob*) and a prepositional phrase (*naar de sterren* 'at the stars') is indicated on the verb's ARG-ST list, cf. (3.20). Since ARG-ST is a lexical feature, it is not included in SYNSEM.

(3.19) Bob kijkt naar de sterren.
Bob looks at the stars
'Bob is looking at the stars.'

(3.20)
$$\begin{bmatrix} \textit{lexeme} \\ \text{PHON} \left\langle \text{kijken} \right\rangle \\ \text{SYNSEM}\,|\,\text{LOC}\,|\,\text{CAT}\,|\,\text{HEAD}\ \textit{verb} \\ \text{ARG-ST} \left\langle \text{NP} \text{ , } \text{PP}\ [\textit{naar}] \right\rangle \end{bmatrix}$$

Argument realization is captured by the valence features, which refer to the syntactic functions of the arguments. A distinction is made between the valence features SUBJ(ECT) and COMP(LEMENT)S.[8] Application to *kijkt* in example (3.19) yields (3.21).

(3.21)
$$\begin{bmatrix} \textit{word} \\ \text{PHON} \left\langle \text{kijkt} \right\rangle \\ \text{SYNSEM}\,|\,\text{LOC}\,|\,\text{CAT} \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{SUBJ} & \left\langle \boxed{1}\ \text{NP} \right\rangle \\ \text{COMPS} & \left\langle \boxed{2}\ \text{PP}[\textit{naar}] \right\rangle \end{bmatrix} \\ \text{ARG-ST} \left\langle \boxed{1} \text{ , } \boxed{2} \right\rangle \end{bmatrix}$$

In contrast to the ARG-ST feature, the valence features are not limited to lexical signs. Instead, they are part of the CATEGORY value, which means they are also assigned to phrasal signs. The relation between argument selection and argument realization is formally captured by the *Argument Realization Principle* (ARP) (Ginzburg & Sag 2000: 23).

[8] Sometimes a third valence feature (SP(ECIFIE)R) is used, but this is not needed if one treats Head-Functor structures as described in (3.17).

(3.22) Argument Realization Principle (ARP):

$$\textit{word} \Rightarrow \begin{bmatrix} \text{SYNSEM}\,|\,\text{LOC}\,|\,\text{CAT} & \begin{bmatrix} \text{SUBJ} & \boxed{A} \\ \text{COMPS} & \boxed{B} \end{bmatrix} \\ \text{ARG-ST}\ \boxed{A} \oplus \boxed{B} & \end{bmatrix}$$

The ARP is a constraint on words. It states that the (possibly empty) list of syntactic arguments is partitioned in two (possibly empty) sublists. The features on the first sublist are realized as subjects (A), whereas the members of the second list are realized as complements (B).

The distinction between the ARG-ST list on the one hand and the valency features on the other hand seems to be redundant at first sight.[9] However, Pollard & Sag (1994: chapter 9), Miller & Sag (1997), Manning & Sag (1999), and Sag (2012) motivate the division into two independent notions for the analysis for a range of phenomena, such as complement extraction, pronominal cliticization, and null instantiation. Those phenomena share the fact that an element that is included in the ARG-ST list, does not appear as a valence feature.

For instance, in the case of complement extraction, as in *Which box did you put the cake in __?*, the extracted complement *which box* appears on the SLASH (or GAP) list and on the ARG-ST list, but it does not appear on the COMPS list (Pollard & Sag 1994: 376–382). In the case of French pronominal affixes, such as *le* 'him' in the example *Marie le voit* 'Marie sees him', the affixes do appear on the ARG-ST list but not on the COMPS list (Miller & Sag 1997: 585–593). In the case null instantiation a lexical sign has undergone a derivation resulting in the omission of a complement, e.g. *eating* in *I'm eating* (Sag 2012: 85–86).

3.2.4 Raising and control

A final concept that will be introduced before turning to the analysis of verb clusters, is the treatment of verbs taking an infinitival complement. A distinction is made between raising and control verbs.

[9] In older versions of HPSG there was no distinction between argument selection and argument realization. The SUBCAT(EGORIZATION) feature was used to capture both phenomena at once (see also section 3.3.2).

Subject and object raising verbs

An example of a *subject raising verb* is *continue* in (3.23).

(3.23) Jack *continues* to visit James.

Jack is both the subject of the verb *visit* and the understood subject of *continue*. The subject of *continue* thus takes on the syntactic and semantic aspects of the subject of the embedded verb. *Subject raising lexemes* (*s-rsg-lx*) are formally defined by the lexical constraint in (3.24).

(3.24) *s-rsg-lx* ⇒ $\left[\text{ARG-ST} \left\langle [\text{LOC } \boxed{1}] , [\text{SUBJ } \langle [\text{LOC } \boxed{1}] \rangle] \right\rangle\right]$

The constraint indicates that *subject raisers* have two syntactic arguments: A subject and an (infinitival) complement. In order to indicate that the subject properties of the raising verbs are shared with the subject of the embedded verb, the LOCAL value of the subject has to be identical to the LOCAL value of the subject of the second argument (1) (Ginzburg & Sag 2000: 22).[10]

The analysis of subject raisers is similar to the treatment of *object raising* or *Accusativus cum Infinitivo* (AcI) verbs, such as *expect* in (3.25).

(3.25) I *expected* John to attend the party.

Those verbs share their object with the subject of the embedded verb, cf. *John* in (3.25). The formal definition of *object raising lexemes* (*o-rsg-lx*) is given in (3.26).

(3.26) *o-rsg-lx* ⇒ $\left[\text{ARG-ST} \left\langle \text{NP} , [\text{LOC } \boxed{1}] , [\text{SUBJ } \langle [\text{LOC } \boxed{1}] \rangle] \right\rangle\right]$

Similar to the subject raising verbs, *structure sharing* is used, but instead of sharing the subjects of both verbs, the LOCAL values of the object are shared with those of the subject of the embedded verb (Ginzburg & Sag 2000: p.22).

Raising verbs do not assign a semantic role to the subject/object that they share with the infinitival complement, i.e. they do not have any requirements on the kind of subject its complement is looking for. As a result, they are compatible with nonreferential subjects, such as *it* in (3.27).

[10] In the raising analysis of Pollard & Sag (1994) the complete *synsem* is shared. Here, only the LOCAL information is shared, thereby allowing for discrepancies between the NONLOCAL values of the raised argument and the unexpressed subject (Ginzburg & Sag 2000: 21, footnote 8).

(3.27) a. It *continues* to rain.

b. I *expected* it to rain.

Another characteristic aspect of raising verbs is the fact that active-passive pairs are semantically equivalent, as shown in (3.28).

(3.28) a. Jack *continues* to visit James.

b. James *continues* to be visited by Jack.

Modal auxiliary verbs are canonically treated as subject raisers in HPSG, see amongst others Warner (2000) and Sag et al. (2003: chapter 13). The Dutch modal and aspectual verbs are typical subject raisers as well, as discussed by Aelbrecht (2009). An example is *moeten* 'have to' in (3.29a). Examples of object raisers are the perception verbs, such as *horen* 'hear' in (3.29b).

(3.29) a. Hij moet daar onmiddellijk mee ophouden!
he has-to there immediately with stop
'He immediately has to stop that!'

b. Ze hoorde hem al komen.
she heard him already come
'She has already heard him coming.'

Subject and object control verbs

While raising verbs do not assign a semantic role to their subject, control verbs, such as *try* and *persuade*, do. Therefore, they are not compatible with nonreferential subjects (3.30), and active constructions are not semantically equivalent to their passive counterparts (3.31).

(3.30) a. * It *tried* to rain.

b. * I *persuaded* it to rain.

(3.31) a. Jack *tried* to see Jill.

b. Jill *tried* to be seen by Jack.

Subject control lexemes (*s-ctrl-lx*), such as *try*, are formally defined by the lexical constraint in (3.32) (Ginzburg & Sag 2000: 22).

(3.32) *s-ctrl-lx* $\Rightarrow$ $\left[\text{ARG-ST} \left\langle \text{NP}_i, [\text{SUBJ} \langle \text{NP}_i \rangle] \right\rangle \right]$

Subject control verbs have two syntactic arguments: A subject and an (infinitival) complement. The fact that those verbs assign a semantic role to their subject is indicated with an index (*i*). This index is shared with the subject NP of the verbal complement.

Since nonreferential NPs have the feature [INDEX *none*], it follows that they are noncompatible with the constraint in (3.32).

Object control lexemes (*o-ctrl-lx*), such as *persuade*, are formally defined by the lexical constraint in (3.33) (Sag et al. 2003: 378).

(3.33) $$\textit{o-ctrl-lx} \Rightarrow \left[\text{ARG-ST} \left\langle \text{NP}, \text{NP}_i, \left[\text{SUBJ} \langle \text{NP}_i \rangle \right] \right\rangle \right]$$

Object control verbs take three syntactic arguments: A subject, an object, and a verbal complement. The index of the object is shared with the subject of the verbal complement.

An example of a Dutch subject control verb is *willen* 'want' in (3.34a). *Vragen* 'ask' in (3.34b) is an example of an object control verb.

(3.34) a. Hij wilde haar verrassen.
he wanted her surprise
'He wanted to surprise her.'

b. Ze vroeg hem wat later te komen.
she asked him little later to come
'She asked him to come a bit later.'

3.2.5 Conclusion

This section has introduced the basic mechanisms of the HPSG framework that will be used in the discussion of the analysis of verb clusters in the remainder of this chapter and in chapters 6 and 7. Besides the use of typed feature structures, the combination of heads and their dependents were described, as well as the syntactic modeling of subcategorization patterns and the analysis of verbs selecting an infinitival complement.

3.3 Generalized raising

This section presents the HPSG analysis of verb clusters that has been widely adopted in the current analyses of the phenomenon, i.e. the argument inheritance mechanism

proposed by Hinrichs & Nakazawa (1994). The original proposal was formulated for German, but it has been employed for the analysis of Dutch verb clusters as well. Before turning to the analysis, the structure of German and Dutch verb clusters will be discussed.

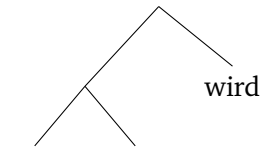
3.3.1 Binary branching verb clusters

Hinrichs & Nakazawa (1994), Kathol (2000), and Müller (2002) assume that German verb clusters have a binary, left-branching structure. Thus, the verb cluster *lesen können wird* 'read can will' in (3.35a) has a structure like (3.36a). Rentier (1994) has applied the German structure to Dutch (see also section 3.4), which results in right-branching trees. For example, the cluster *zal kunnen lezen* 'will can read' in (3.35b) is assigned the structure in (3.36b).

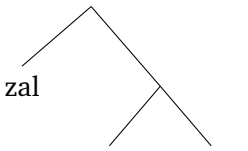
(3.35) a. weil Ulrich dieses Buch *lesen können wird*.
because Ulrich this book read can will

b. omdat Ulrich dit boek *zal kunnen lezen*.
because Ulrich this book will can read
'because Ulrich will be able to read this book.'

(3.36) a.



b.



The motivation for left-branching structures in German is found in the *Oberfeldumstellung* or *Auxiliary Flip*, and the (partial) fronting of verb clusters. As mentioned in chapter 1 (section 1.11), German clustering verbs generally occur to the right of the verb they select, as in (3.37a) (Kathol 2000: 180). *Oberfeldumstellung* refers to constructions in which the finite verb is the leftmost verb in the cluster, such as *wird* 'will' in (3.37b) (Kathol 2000: 187). Such constructions can be elegantly dealt with if they are assigned a structure in which the non-finite verbs form a unit.

(3.37) a. dass Peter das Buch *finden*$_3$ *können*$_2$ *wird*$_1$.
that Peter the book find can will

b. dass Peter das Buch *wird*$_1$ *finden*$_3$ *können*$_2$.
that Peter the book will find can
'that Peter will be able to find the book.'

The second motivation is (partial) VP fronting. The examples in (3.38) illustrate that not only phrasal constituents can be fronted such as the NP *das Buch* in (3.38a). Verbs can be fronted as well, either alone (3.38b) or as a cluster (3.38c) (Kathol 2000: 186). The fronted cluster *finden können* 'find can' in (3.38c) is similar to the subcluster in the left-branching structure in (3.36a).

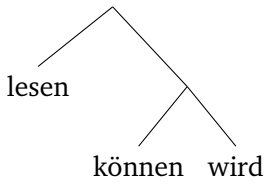
(3.38) a. Das Buch wird Peter finden können.
the book will Peter find can

b. *Finden* wird Peter das Buch können.
find will Peter the book can

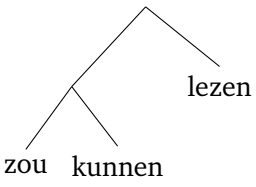
c. *Finden können* wird Peter das Buch.
find can will Peter the book
'Peter will be able to find the book.'

Kiss (1994, 1995) proposes a right-branching structure for German, cf. (3.39a). According to Kiss, verbs that cluster obligatorily (such as *werden* 'will' and *können* 'can') can only select lexical items. For clusters containing multiple obligatorily coherent verbs this implies that the structure in (3.39a) is the only one possible: *wird* first selects the lexical element *können* which selects in its turn the lexical verb *lesen*. This contrasts with the structure in (3.36a), in which *wird* selects the verbal complex *lesen können* (Kiss 1994: 92-96). The tree in (3.39b) presents the corresponding left-branching structure for Dutch.

(3.39) a.



b.



The right-branching analysis for German is disfavoured by several authors, since such tree structures cannot adequately explain *Oberfeldumstellung* and *partial verb fronting*. For a discussion, see Kathol (2000: 183–187) and Müller (2002: 111–112).

Kiss admits that his right-branching analysis cannot account for the Oberfeldumstellung, but adds that it can easily deal with the *Zwischenstellung*, i.e. constructions in which the finite verb appears in between the other verbs of the verb cluster (Kiss 1994: 98, fn).

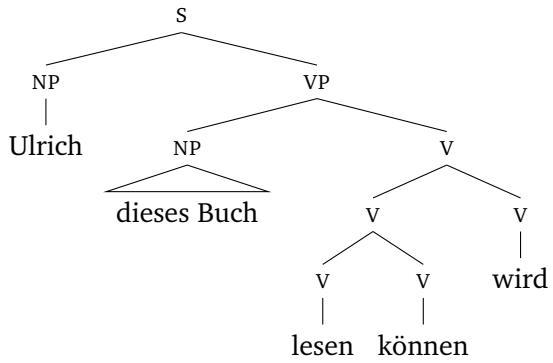
(3.40) dass Peter das Buch $finden_3$ $wird_1$ $können_2$. (Zwischenstellung)
that Peter the book find will can
'that Peter will be able to find the book.'

Kiss argues that neither his nor the left-branching analyses can cope with the full range of data: While his right-branching analysis cannot deal with the Oberfeldumstellung, the Zwischenstellung is less trivial to account for in a left-branching analysis. However, the Oberfeldumstellung is accepted by almost all speakers of German, while the Zwischenstellung is not (Wurmbrand 2004: 50). This pleads in favour of the left-branching analysis, in addition to the fact that it covers more constructions (i.e. the Oberfeldumstellung and partial verb fronting).

3.3.2 Argument Inheritance

Hinrichs & Nakazawa (1989, 1994) argue that in constructions with verb clusters, main verbs combine with auxiliaries before combining with their non-verbal complements. They opt for a left-branching structure to which the arguments of the main verb are added one at a time, as in (3.41).

(3.41)

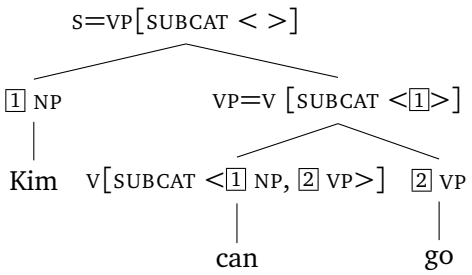


This proposal differs from previous analyses of the German VP structure, as proposed by amongst others den Besten & Edmondson (1983) and Uszkoreit (1987a,b), in which main verbs first combine with their arguments before being selected by other verbs (Hinrichs & Nakazawa 1989: 193–194). However, a similar analysis can be found in both Steedman's functional categorial treatment and in Johnson's GPSG analysis (see section 3.1).

Also within HPSG the Hinrichs-Nakazawa analysis is not completely new, since it is an extension of an existing HPSG principle: The *Subcategorization Principle*. It dates from a time when only one feature was used to specify valence, i.e. SUBCAT(EGORIZATION). The principle states that "in a *headed phrase* (...), the SUBCAT value of the head daughter is the concatenation of the phrase's SUBCAT list with the list (in order of increasing obliqueness) of SYNSEM values of the complement daughters" (Pol-

lard & Sag 1994: 34). The tree in (3.42) shows an example of the subcategorization principle for the sentence *Kim can go* (Pollard & Sag 1994: 42).

(3.42)



In (3.42), the SUBCAT value (<1NP,2VP>) of *can* is the concatenation of the SUBCAT value of its mother (<1>) and its verbal complement (<2>). The effect of the Subcategoration Principle is thus to 'check off' the subcategorization requirements of a head as they become satisfied by its dependents.

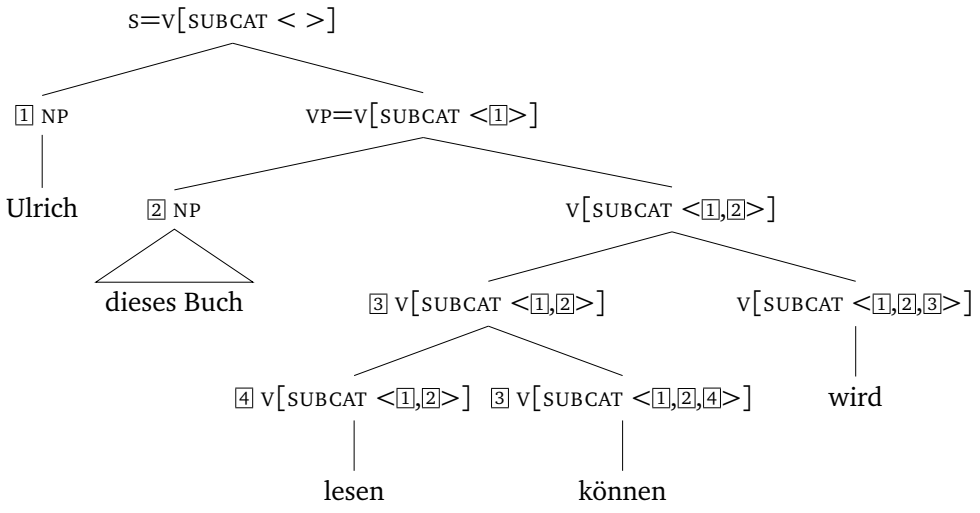
Hinrichs & Nakazawa (1994) point out that, contrary to the English auxiliaries, which have a subject and a verbal complement in their SUBCAT list, the SUBCAT list of German auxiliaries not only consists of an infinitive, but of the verbal complement AND the complement's SUBCAT list. The valence restrictions are formally described in (3.43), after Hinrichs & Nakazawa (1994).

(3.43)

$$\left[\text{SUBCAT } \boxed{A} \oplus \left\langle \left[\text{LOCAL}\,|\,\text{CAT} \left[\begin{array}{ll}\text{HEAD} & \textit{verb}\\ \text{SUBCAT} & \boxed{A}\end{array}\right]\right]\right\rangle\right]$$

Adding this information to (3.41), yields (3.44).

(3.44)



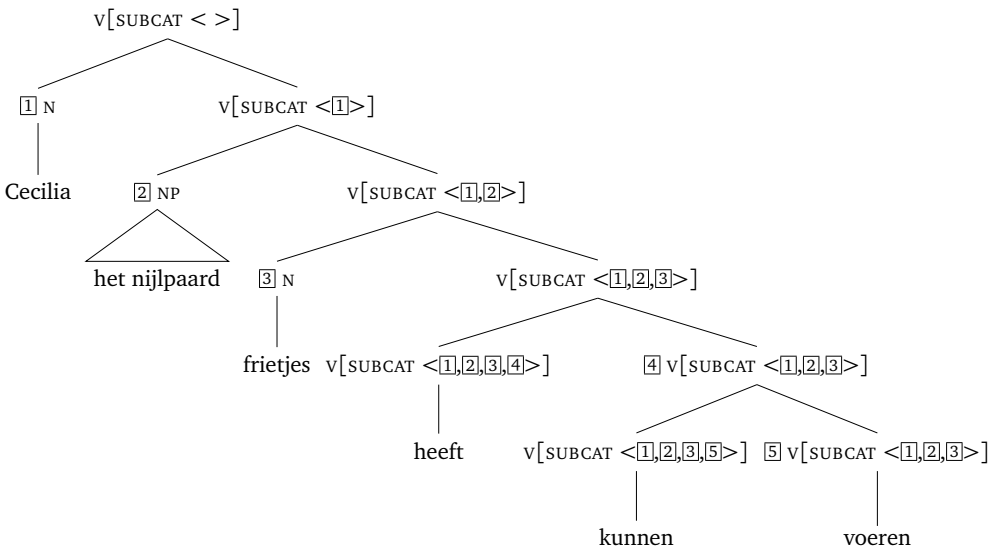
The Hinrichs-Nakazawa analysis extends the Subcategorization Principle to a more general *Argument Inheritance* model which also takes into account the raising of non-subject arguments.

Verbs which select an infinitival complement such as *können* 'can' in (3.44), inherit the selection restrictions of that complement (i.e. the NP complements [1] and [2]) and the infinitival complement itself ([4]). The unsaturated complements are thus passed on to the mother node via the selecting verb. The motivation for doing this is that the Subcategorization Principle needs no adaptations (Hinrichs & Nakazawa 1994: 22). The argument inheritance mechanism also applies to *wird*, which selects *lesen können* ([3]). As it inherits an unsaturated verb cluster, it also inherits the unsaturated NP complements ([1] and [2]). Towards the top of the tree, the NP complements are cancelled off in the regular way.

The tree in (3.46) illustrates the argument inheritance principle applied to the Dutch construction in (3.45).

(3.45) ... dat Cecilia het nijlpaard frietjes heeft kunnen voeren.
... that Cecilia the hippo chips.DIM has can feed
'that Cecilia was able to feed the hippo chips.'

(3.46)



In (3.45) the verb *kunnen* 'can' selects the infinitive *voeren* 'feed' ([5]), while inheriting the valence restrictions of that verb as well: A subject that feeds (in this case the NP *Cecilia* [1]), an object that gets fed (the NP *het nijlpaard* 'the hippo' [2]), and what it is fed (the NP *frietjes* 'chips' [3]). The verb *heeft* 'has' selects *kunnen voeren* 'can feed' ([4]), and inherits the valence requirements of its complement. So, the verb cluster *heeft*

kunnen voeren 'has can feed' as a whole selects the arguments of the main verb *voeren* 'feed'. Similar to the the German example, the NP complements are cancelled off in the regular way towards the top of the tree.

In more recent versions of HPSG the SUBCAT list is replaced by the syntactic valence features SUBJ(ECT) and COMP(LEMENT)S on the one hand (Pollard & Sag 1994: chapter 9), and the ARG(UMENT)-ST(RUCTURE) list on the other hand (Miller & Sag 1997) (see section 3.2.3).[11] The valence restrictions for auxiliaries in (3.43) can be reformulated as (3.47).[12]

(3.47)
$$\left[\begin{array}{l}\text{ARG-ST}\ \left\langle \boxed{1} \right\rangle \oplus \boxed{A} \oplus \left\langle \left[\begin{array}{l}\text{LOCAL}\,|\,\text{CAT}\ \left[\begin{array}{ll}\text{HEAD} & \textit{verb}\\ \text{SUBJ} & \left\langle \boxed{1} \right\rangle\\ \text{COMPS} & \boxed{A}\end{array}\right]\end{array}\right] \right\rangle\end{array}\right]$$

If the COMPS list ($\boxed{A}$) is empty, the constraint is similar to the subject raising constraint, see (3.24) in section 3.2.4. Therefore argument inheritance is also known as *generalized raising*, since it extends the subject raising mechanism to all arguments of the verb.[13]

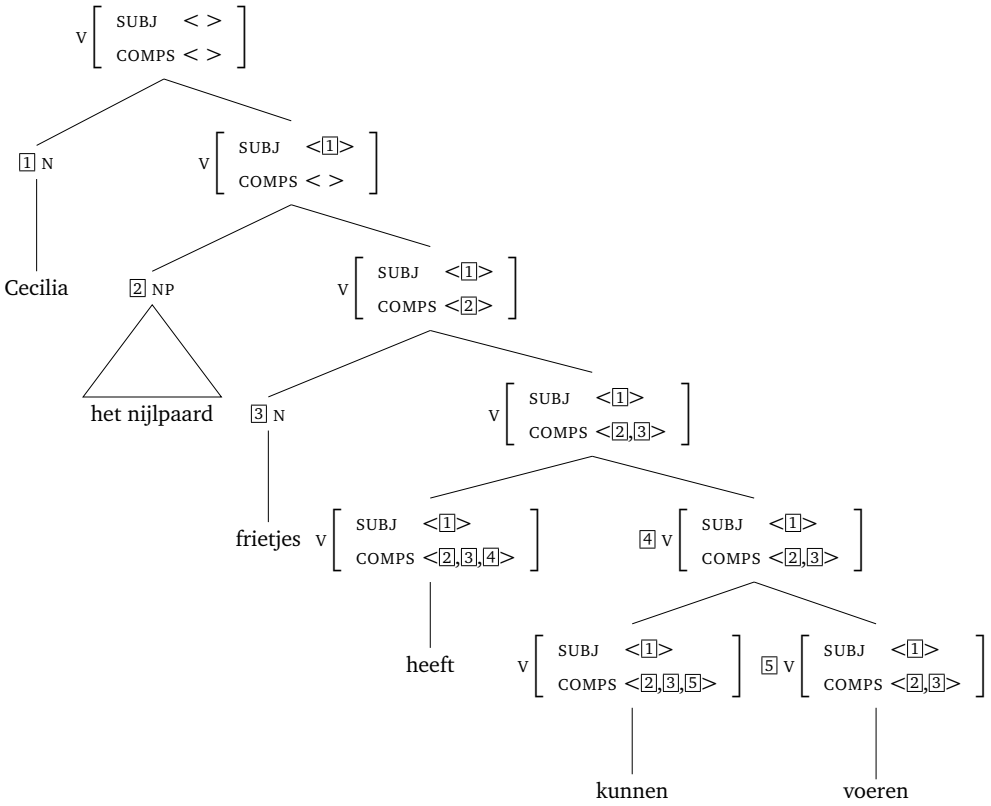
Application of (3.47) to the sentence in (3.45) yields a tree structure like (3.48).

[11] The third valence feature (SP(ECIFIE)R), which is mainly used to model the selection of a determiner by a nominal, is omitted here as it is not relevant in the treatment of argument inheritance. In order to account for analyses that make use of multiple valence features instead of one SUBCAT feature, the Subcategorization Principle is reformulated as the Valence Principle (Pollard & Sag 1994: 348).

[12] Although Hinrichs & Nakazawa (1994) do not make use of SUBJ and COMPS, they are used by several authors that have adapted the argument inheritance analysis (sections 3.4 and 3.5). The conversion to an analysis with multiple valence features is presented here in order to compare the analyses discussed in the remainder of this chapter.

[13] Languages that do not allow complement raising, such as English, obligatorily have an empty COMPS list for phrases. Those languages are subject to the Empty COMPS Constraint (ECC), cf. Ginzburg & Sag (2000: 33).

(3.48)



The Hinrichs-Nakazawa analysis was originally formulated for German, but as (3.46) and (3.48) show, the principle can be applied to Dutch as well. The analysis is adopted by amongst others Rentier (1994), Bouma & van Noord (1998), Kathol (2000), and Müller (2002).

Besides the analysis of verb clusters, argument inheritance was also adopted for the treatment of clitic climbing in the Romance languages, e.g. by Abeillé et al. (1998) for French and by Monachesi (1998) for Italian.

3.4 Head-Complement versus Head-Cluster structures

This section compares some of the applications of argument inheritance for the analysis of (German and Dutch) verb clusters. According to Hinrichs & Nakazawa (1994), the verbs within the cluster combine with their complement in the same way as heads combine with their complements outside the cluster. This contrasts with the other approaches presented in this section, i.e. the analyses of Rentier (1994), Kathol (2000), and Müller (2002). They distinguish cluster formation from the combinatorics used

outside the cluster.

Before turning to the theory, two notational issues should be pointed out. First, some authors make use of the older SUBCAT feature to describe a verb's valence, while others differentiate between SUBJ and COMPS. Second, several authors make use of *immediate dominance* (ID) rules and *linear precedence* (LP) constraints to describe the hierarchy and the linear order of phrase structure rules respectively. While traditional phrase structure rules such as S → NP VP capture both dominance and linear order, the ID rule S → NP, VP states that an the S node immediately dominates an NP and a VP node, without saying something about the linear order of the NP and VP nodes. The linear precedence constraint NP < VP states that the NP should precede the VP node. The idea to distinguish between dominance and ordering originated within GPSG, but was taken over in other frameworks, such as HPSG.

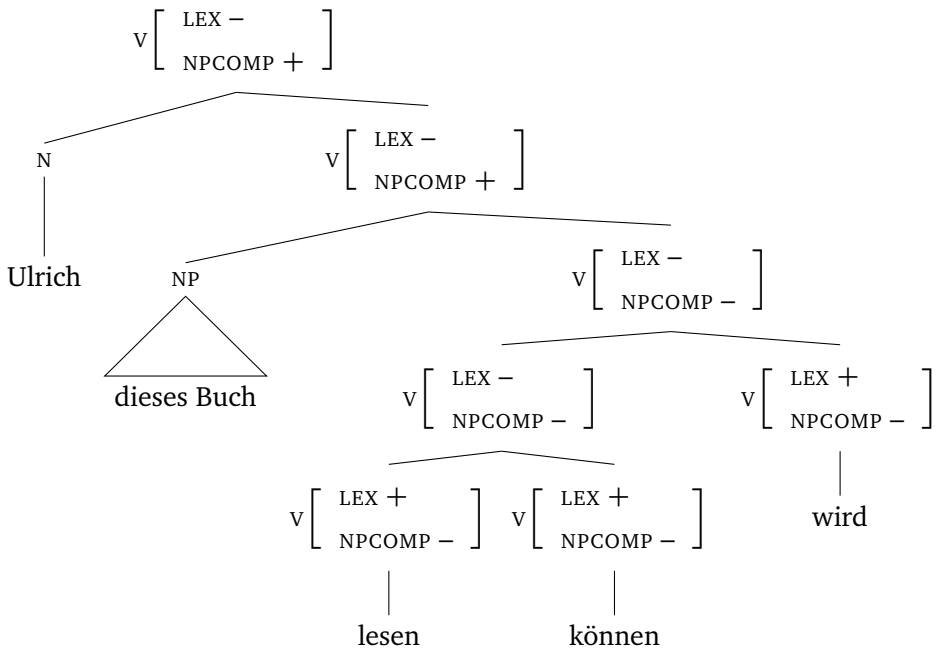
3.4.1 Hinrichs & Nakazawa (1994)

Hinrichs & Nakazawa (1990, 1994) treat verb clusters in the same way as other constituents, i.e. they consider the combination of main verbs and auxiliaries as regular head-complement structures (Hinrichs & Nakazawa 1990: 13).[14] In order to ensure that verb clusters do not contain any NP complements, they introduce the binary NPCOMP feature. A negative NPCOMP feature indicates that there are no NPs included in the constituent, while a positive value is assigned to constituents containing NP complements. Lexical verbs are lexically specified as [LEX +][NPCOMP –], verbal complexes are marked [LEX –][NPCOMP –], and larger constituents consisting of a verb cluster and non-verbal arguments are marked [LEX –][NPCOMP +], cf. the tree in (3.49) (Hinrichs & Nakazawa 1994: 22–23).[15]

[14] Bouma & van Noord (1998) do not make a distinction between Head-Complement and Head-Cluster structures either, but since their analysis differs from the canonical HPSG approaches, it is presented in a separate section (section 3.5).

[15] NPCOMP and LEX are syntactic features directly embedded under CAT. They are neither valence features like SUBCAT nor head features.

(3.49)



The following ID rules state that verb clusters have a negative NPCOMP feature (3.50a) while the NPCOMP feature of constituents with NP complements is positive (3.50b):[16]

(3.50) a. V[NPCOMP –] → H[LEX +], V

b. V[NPCOMP +] → NP, H

The introduction of the NPCOMP feature makes it possible to extend the analysis of constructions with verb clusters, like the tree representation in (3.49) or the sentence in (3.51a) to the analysis of constructions with *cluster creepers*, i.e. constructions in which non-verbal material interrupts the verb cluster.[17] An example is given in (3.51b) (Hinrichs & Nakazawa 1994: 27–33).[18]

(3.51) a. . . . dass du uns die Schlacht *hast gewinnen helfen*.
. . . that you us the battle have win help

b. . . . dass du uns *hast* die Schlacht *gewinnen helfen*.
. . . that you us have the battle win help
'. . . that you helped us to win the battle.'

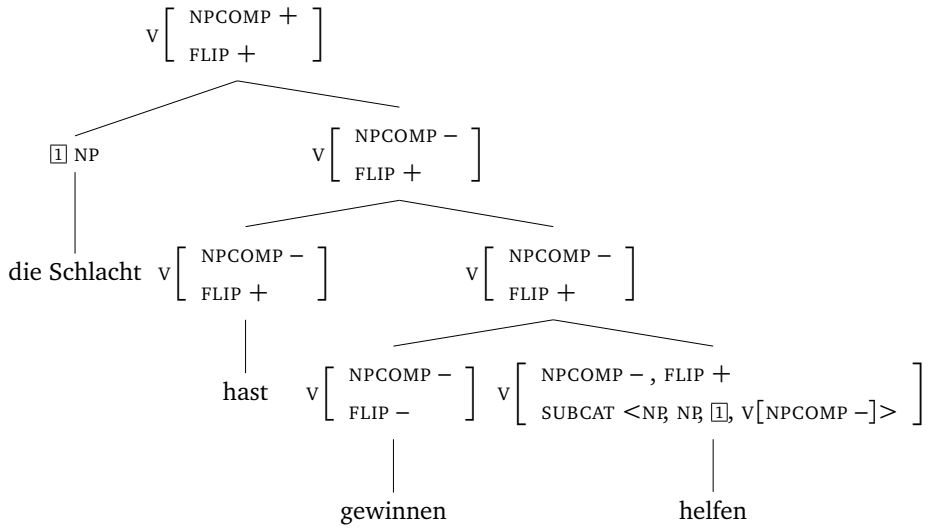
[16] H stands for HEAD.

[17] Such constructions are known as *VP raising* in transformational grammar (cf. section 2.2).

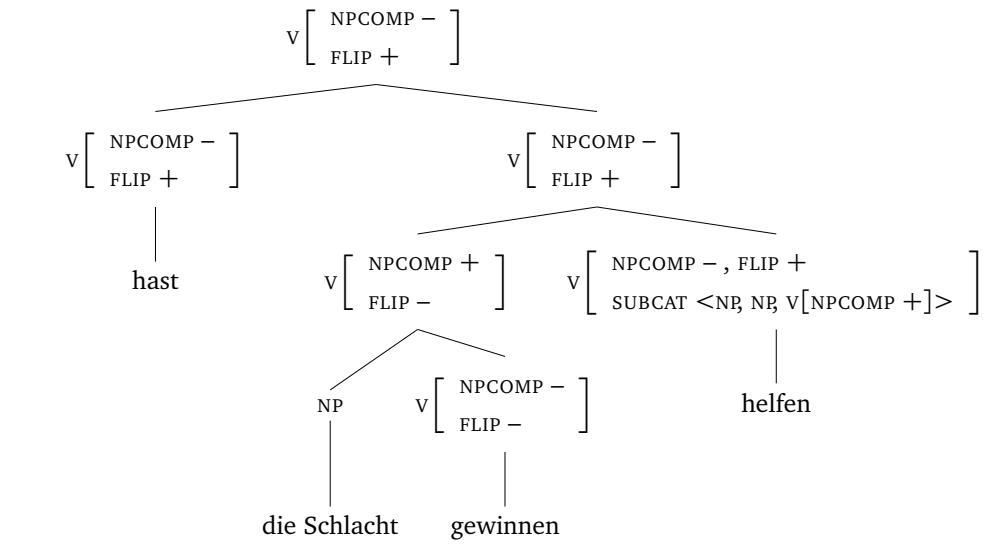
[18] Hinrichs & Nakazawa (1994) describe sentences like (3.51b) as constructions in which the auxiliary has flipped over the verb cluster and at least one NP complement, instead of only flipping over the verb cluster.

Hinrichs & Nakazawa (1994) argue that the difference between constructions like (3.51a) and (3.51b) can be specified in the SUBCAT feature of the selecting verbs. For example, in the case of regular verb cluster formation verbs like *helfen* 'help' and *lassen* 'let' select a lexical complement, i.e. a verbal complement without NPs, indicated by the feature V[NPCOMP –]. In the case of VP-raising, a complement is selected that does include NPs (V[NPCOMP +]). The difference between both constructions is illustrated in the tree structures in (3.52) (Hinrichs & Nakazawa 1994: 28). Note also that in both constructions in (3.51) *helfen* 'help' appears as IPP and triggers *auxiliary flip*, i.e. the auxiliary *hast* 'have' occurs as the first verb in the cluster. In order to model the auxiliary flip, Hinrichs and Nakazawa employ the binary head feature FLIP. The positive value of this feature appears in the lexical categories of the triggering IPP verb and the tense auxiliary, from which it is propagated to the mother nodes (Hinrichs & Nakazawa 1994: 24–27).

(3.52) a.



b.



In (3.52a), *helfen* 'help' only selects *gewinnen* 'win', which is [NPCOMP –], while in (3.52b) it selects the VP *die Schlacht gewinnen* 'win the battle', which is [NPCOMP +]. Hinrichs & Nakazawa (1994) thus propose two lexical entries for *helfen*: One in which it subcategorises for an [NPCOMP –] complement, and one in which it subcategorises for an [NPCOMP +] complement. The selecting verb *helfen* itself has the feature [NPCOMP –] in both cases, as it is still unsaturated for the NP complements *du* 'you' and *uns* 'us'. This is also the case for the mother node of *helfen* in both (3.52a) and (3.52b).

Since the NPCOMP feature is rather arbitrary, several authors abandon it and replace it by one or more other features. Moreover, it cannot account for constructions in which the verb cluster is interrupted by other elements than NPs, such as APs or PPs. The authors admit that for some speakers the restrictions on the type of complements that appear within the cluster should be less constrained (Hinrichs & Nakazawa 1994: 33–35).

3.4.2 Rentier (1994)

Rentier (1994) extends Hinrichs and Nakazawa's argument inheritance analysis to Dutch. Furthermore, he adds some refinements, such as the distinction between a *Phrase Structure schema* for the formation of regular head-complement phrases, and a *Cluster Formation schema* for the formation of constructions with a verb cluster. They are captured by the ID statements in (3.53), making use of the features H(EAD),

S(UBJ), and C(OMPS). Rentier assumes a flat clause structure for Dutch, but verb clusters are strictly binary branching.

(3.53) a. Phrase Structure schema
$XP_{[LEX-]} \rightarrow S, C_1, .., C_n, H_{[GOV<>, LEX+]}$

b. Cluster Formation schema
$X_{[LEX+]} \rightarrow H_{[GOV<C_i>, LEX+]}, C_i$

A remarkable adaptation concerns the use of the LEX feature. Hinrichs and Nakazawa assign the value [LEX +] to all lexical verbs and [LEX –] to verbal complexes consisting of a main verb and auxiliaries (Hinrichs & Nakazawa 1994: 22). In Rentier's treatment of verb clusters, however, the *Cluster Formation schema* in (3.53b) gives rise to complex words that, just like terminal nodes, receive the feature LEX+ (Rentier 1994: 819). The adaptation of the LEX feature makes the NPCOMPS feature, introduced by Hinrichs & Nakazawa (1994), superfluous.

Rentier furthermore adds a GOV(ERNOR) feature to the head of the cluster. GOV is a valence feature that governs one complement, and allows to define the difference between Dutch and German verb clusters. This difference can be captured by the LP constraints in (3.54). Rentier argues that the matrix verb (or *governing verb*) occurs leftmost in Dutch verb clusters (3.54a), while in German the matrix verb is the rightmost verb in the cluster (3.54b) (Rentier 1994: 820).

(3.54) a. Dutch:
[GOV <X>] < X

b. German:
X < [GOV <X>]

The distinction between verb clusters and regular head-complement structures is adopted by several others, such as Kathol (2000) and Müller (2002).

3.4.3 Kathol (2000) and Müller (2002)

Both Kathol (2000) and Müller (2002) argue that the selection of the verbal complement had better not be integrated in a verb's SUBCAT list, since verb clusters can be subject to (partial) fronting (see section 3.3.1). Therefore, Kathol introduces VCOMPL, a separate valence feature for verbal complements, which is analogous to Rentier's GOV feature (Kathol 2000: 188–189). Verbs in coherent (or clustering) constructions

form a *verbal complex* with their infinitival complement via VCOMPL (3.55a), whereas verbs in incoherent (or non-clustering) constructions do not. In incoherent constructions, the verbal complement appears on the verb's SUBCAT list, as in (3.55b) (Kathol 2000: 193).

(3.55) a. coherent construction

$$\left[\begin{array}{l}\text{VAL}\left[\begin{array}{l}\text{SUBCAT } \boxed{1}\\ \text{VCOMPL}\left\langle\left[\text{VAL}\left[\begin{array}{l}\text{SUBCAT } \boxed{1}\\ \text{VCOMPL } \langle\,\rangle\end{array}\right]\right]\right\rangle\end{array}\right]\end{array}\right]$$

b. incoherent construction

$$\left[\begin{array}{l}\text{VAL}\left[\begin{array}{l}\text{SUBCAT}\left\langle \boxed{1}, \left[\begin{array}{l}\text{SUBCAT } \boxed{1}\\ \text{VCOMPL } \langle\,\rangle\end{array}\right]\right\rangle\\ \text{VCOMPL } \langle\,\rangle\end{array}\right]\end{array}\right]$$

Constructions that cluster obligatorily, such as *scheinen* 'seem, appear' in (3.56a–b), can thus only be licensed by the schema in (3.55a), whereas optionally coherent verbs such as *versuchen* 'try' in (3.56c–d) can be licensed by both (3.55a) and (3.55b).

(3.56) a. weil Lisa zu schlafen *scheint*.
because Lisa to sleep seems

b. * weil Lisa *scheint* zu schlafen.
because Lisa seems to sleep
'because Lisa seems to sleep.'

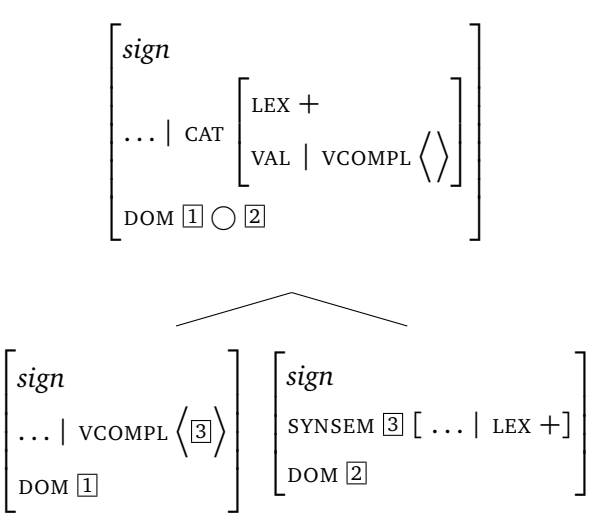
c. weil Lisa zu schlafen *versucht*. (coherent)
because Lisa to sleep tries

d. weil Lisa *versucht* zu schlafen. (incoherent)
because Lisa tries to sleep
'because Lisa tries to sleep.'

Kathol replaces Hinrichs and Nakazawa's FLIP feature by G(O)V(ERN)OR in order to model the word order of the verbs in the cluster. If a verb has the feature [GVOR →], its governor should appear to its right, while the governor of verbs with the feature [GVOR ←] should appear to the left (e.g. in the case of Oberfeldumstellung in German). GVOR can be underspecified as [GVOR *dir*], in order to deal with verbs that can select their complement to the left or to the right, e.g. Dutch *wil lezen* versus *lezen wil* 'wants to read' (Kathol 2000: 199–200).

The accounts of word order so far are simplified versions of the facts, as they only deal with canonical word order and the Oberfeldumstellung. As discussed in section 3.3.1, instances of the Zwischenstellung are much harder to deal with in a left-branching analysis of German. In order to account for such constructions, Kathol does not deal with word order by means of phrase structure trees, but employs an additional level of representation to model the linear representation of a construction, i.e. the *domain* level, following Reape (1993). Each sign is assigned a DOM(AIN) feature, which has a list of elements of type *sign* as its value. In head-VCOMPL constructions, the domains of both daughters are merged (indicated by ◯), as illustrated in (3.57) (Kathol 2000: 209):

(3.57)



By means of linearization constraints, the elements in the DOM feature of the top of the verbal complex can be shuffled. For the Zwischenstellung in (3.40), repeated in (3.58), the DOM feature of the verb cluster is presented in (3.59) (Kathol 2000: 210).

(3.58) dass Peter das Buch *finden*$_3$ *wird*$_1$ *können*$_2$.
that Peter the book find will can
'that Peter will be able to find the book.'

(3.59)
$$\left[\text{DOM} \left\langle \left[\langle \textit{finden} \rangle \; \text{GVOR} \rightarrow\right], \left[\langle \textit{wird} \rangle\right], \left[\langle \textit{können} \rangle \; \text{GVOR} \leftarrow\right] \right\rangle\right]$$

While Kathol is able to account for non-canonical word orders such as the Oberstellung and the Zwischenstellung, he gives up the surface-oriented analysis in order to model word order in German and Dutch.

Kathol allows separable verb particles in VCOMPL for the analysis of German verb clusters in general (Kathol 2000: 191), but for the analysis of genuine VP-raising, i.e. in order to include *cluster creeping* of non-verbal complements, Kathol further refines the analysis of Hinrichs & Nakazawa (1994). By means of the lexical rule (LR) in (3.60) Kathol allows non-verbal elements to enter the verb cluster, which is indicated by the feature [LEX –]. The rule is only applicable to verbs allowing Oberfeldumstellung, as indicated by [GVOR ←] (Kathol 2000: 229–231).

(3.60)
$$\left[\begin{array}{l}\text{HEAD}\left[\begin{array}{l}\text{GVOR}\ \leftarrow\\ \text{VFORM}\ \textit{nfin}\end{array}\right]\\ \text{VCOMPL}\left\langle\left[\begin{array}{l}\text{LEX}\ +\\ \text{SUBCAT}\ \boxed{1}\end{array}\right]\right\rangle\\ \text{SUBCAT}\ \boxed{1}\end{array}\right] \Longrightarrow_{LR} \left[\begin{array}{l}\text{HEAD}\left[\begin{array}{l}\text{GVOR}\ \leftarrow\\ \text{VFORM}\ \textit{nfin}\end{array}\right]\\ \text{VCOMPL}\left\langle\left[\begin{array}{l}\text{LEX}\ -\\ \text{SUBCAT}\ \boxed{2}\end{array}\right]\right\rangle\\ \text{SUBCAT}\ \boxed{2}\end{array}\right]$$

The SUBCAT list in the VP-raising construction (2) is shorter in comparison to the SUBCAT list in the case of a verb cluster without interrupting elements, as the verbal element on VCOMPL is partially saturated. In both cases SUBCAT list is structure-shared between the governing verb and the verbal complement.

For the analysis of the third construction Kathol argues that, similar to Dutch, the third construction in German (3.61c) seems to stand midway between coherent (3.61a) and incoherent structures (3.61b) (Kathol 2000: 243).

(3.61) a. dass er es zu reparieren versucht.
that he it to repair tries

b. dass er versucht [es zu reparieren].
that he tries it to repair

c. dass er **es** versucht [zu reparieren].
that he it tries to repair
'that he tries to repair it.'

Kathol follows den Besten & Rutten (1989) by analysing the third construction as an extraposed construction of which (some of) the arguments are realised in the Mittelfeld. The SUBCAT requirements of the selecting verbs can be formally defined as follows:[19]

[19] In HPSG-based analyses of German, it is common to use one SUBCAT feature for all valence elements of finite verbs. In the case of non-finite verbs, however, a separate SUBJ feature is employed for the subject requirements of non-finite verbs. (Kathol 2000: 16).

(3.62)
$$\begin{bmatrix} \text{SUBCAT } \boxed{2} \oplus \left\langle \text{V} \begin{bmatrix} \text{SUBCAT } \boxed{1} \oplus \boxed{2} \\ \text{SUBJ } \boxed{1} \end{bmatrix} \right\rangle \\ \text{VCOMPL} \left\langle \ \right\rangle \end{bmatrix}$$

The verbal complement is selected via the SUBCAT list of the matrix verb, while VCOMPL is empty. Moreover, some or even all of the non-subject arguments (2) are raised to the SUBCAT list of the matrix verb (Kathol 2000: 248–249).

Müller (2002) models the formation of verb clusters in a similar way as Kathol. He makes use of a specific *cluster* feature XCOMP in order to define verb clusters.[20] In addition, he uses the LEX feature in a similar fashion as Rentier, i.e. a verb cluster receives the feature [LEX +].[21] Müller's *cluster schema* is given in (3.63).

(3.63)
$$\textit{head-cluster-structure} \Rightarrow \begin{bmatrix} \text{SYNSEM} \begin{bmatrix} \text{LOC} \mid \text{CAT} \mid \text{XCOMP } \boxed{1} \\ \text{LEX} + \end{bmatrix} \\ \text{HEAD-DTR} \left[\text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{XCOMP } \boxed{1} \oplus \left\langle \boxed{2} \right\rangle \right] \\ \text{NON-HEAD-DTRS} \left\langle \left[\text{SYNSEM } \boxed{2} \right] \right\rangle \end{bmatrix}$$

The cluster schema states that the head daughter (HEAD-DTR) selects a complement via XCOMP (2). The remainder (if any) of the XCOMP list (1) is passed on the mother node.[22] In contrast to verb clusters, regular head-complement structures have the feature LEX – and they have an empty XCOMP list. Similar to Kathol's VCOMP feature, XCOMP may contain separable verb particles (Müller 2002: 86–88).

3.5 An alternative analysis: flat tree structures

In contrast to the binary-branching tree structures that were advocated in the previous sections, Bouma & van Noord (1998) use 'flat' tree structures for verb clusters, as

[20] In this context XCOMP is similar to Kathol's VCOMPL and Rentier's GOV feature, but contrary to VCOMPL and GOV the value of XCOMP is not limited to verbs. It can for example contain predicative complements as well (Müller 2002: 86).

[21] In Müller's analysis, [LEX –] is assigned if a head is combined with its complement; otherwise the construction is [LEX +]. If an unsaturated verb is combined with an adjunct it is still [LEX +] (Müller 2002: 87).

[22] In most cases 1 is the empty list, but in constructions with particle verbs it contains the particle, e.g. *an* in *Es fing zu regnen an* 'It started to rain' (Müller 2002: 87).

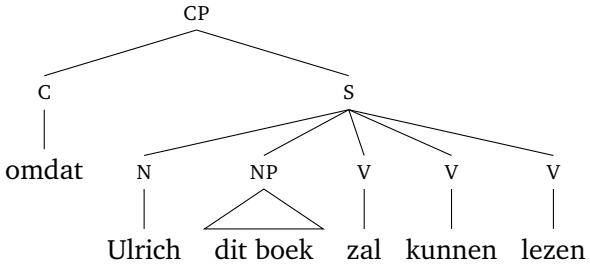
illustrated in (3.64) for the constructions in (3.35).

(3.64) a.

b.

In fact, they assign flat structures to entire clauses, which implies that non-verbal elements are sister nodes of the verbs, as in (3.65) for the construction in (3.35b).

(3.65)



Similar to Hinrichs & Nakazawa (1994), Bouma & van Noord (1998) use only one head-complement schema, i.e. they do not use a specific valence feature to mark the verb cluster. The major difference with the binary-branching clusters is that the verbs combine with both the infinitival complement and the (non-verbal) complements of that complement in one step (Bouma & van Noord 1998: 43).

An important argument for this analysis is the fact that it is hard to deal with all possible word orders in a binary branching structure. Consider, for instance, the *Zwischenstellung*, as exemplified in (3.66). In that example the verb *hätte* 'should have' does not occur as the leftmost verb in the cluster; it has only 'flipped' over the modal *müssen* 'must'. Left-branching structures can deal with the Oberfeldumstellung, such as example (3.37) in section 3.3.1, but not with the Zwischenstellung (Bouma & van Noord 1998: 50).

(3.66) zu dem Zeitpunkt an dem ich mich *entscheiden*$_3$ *hätte*$_1$ *müssen*$_2$
at the point at which I me decide had must
'at the point at which I shoud have made a decision'

In order to explain such constructions, as well as canonical verb cluster constructions, Bouma & van Noord (1998) use only one head-complement schema, which also includes the modifiers of the verbs. It is combined with linear precedence statements to account for the correct word order. Those linear precedence rules are determined by three principles: Directionality, Topology, and Government.

Directionality This principle determines the position of the complement daughter(s) relative to the head daughter. If a complement has the feature [DIR →], it must precede the head. If it has the feature [DIR ←], the complement follows the head (Bouma & van Noord 1998: 52).

Topology Topological constraints determine the position of the verb cluster relative to the position of the non-verbal elements within the VP. A distinction is made between only two topological fields: An *inner zone* and an *outer zone*. The inner zone contains the verb cluster, whereas the outer zone contains the non-verbal arguments (Bouma & van Noord 1998: 53–55). The idea of integrating topological fields into HPSG is also present in Kathol (2000), although he makes use of the traditional topological fields.

Government This principle determines word order within the verb cluster. This is modeled by means of the GVOR feature, following Kathol (2000), cf. section 3.4.3. For instance, if a verb has the feature [GVOR →], its governor must appear to the right (Bouma & van Noord 1998: 55–60).

As an example of how the principles interact, consider the construction in (3.67):

(3.67) dat Jan Marie het boek wil laten lezen.
that Jan Marie the book wants let read
'that Jan wants to let Marie read the book.'

The lexical entry for the object raising verb *laten* 'let' is given in (3.68), showing how the three principles of directionality, topology and government are combined.

(3.68)
$$\left\langle \text{laten}, \left[\begin{array}{l} \text{HEAD } \textit{verb}[\textit{inf}] \\ \text{SUBJ } \left\langle \boxed{2}\ \text{NP} \right\rangle \\ \text{COMPS } \left\langle \boxed{3}\ \text{NP}[\textit{acc}, \text{DIR} \rightarrow] \right\rangle \oplus \boxed{1} \oplus \left\langle \boxed{4} \left[\begin{array}{l} \text{HEAD } \textit{verb}[\textit{inf}] \\ \text{COMPS } \boxed{1} \\ \text{ZONE } \textit{inner} \\ \text{GVOR } \leftarrow \end{array}\right] \right\rangle \\ \text{ARG-S } \left\langle \boxed{2},\boxed{3},\boxed{4} \right\rangle \end{array}\right] \right\rangle$$

Since *laten* selects a verbal complement that is realised within the verb cluster, the ZONE value of its complement is *inner*. The verbal complement occurs to the right of

laten, so it has a [GVOR ←] feature, whereas the NP complement has to occur to the left of *laten*, so it has a [DIR →] feature (Bouma & van Noord 1998: 64).

The advantage of the analysis proposed by Bouma & van Noord (1998) is that they do not need a cluster formation schema to analyse verb clusters and that they can deal with the interruption of non-verbal elements within the verb cluster without an NPCOMP feature, but the downside of this approach is that more features and more complex word order constraints need to be introduced in order to avoid overgeneration.

3.6 Conclusion

This chapter provided an overview of the current monostratal treatments of verb clusters in German and Dutch. After giving a brief sketch of how verb clusters are analysed within CG and GPSG, the HPSG framework was introduced, as it will be the main framework of analysis in the remainder of this thesis.

Furthermore, the *argument inheritance* or *generalized raising* mechanism as proposed in Hinrichs & Nakazawa (1994) was presented, as it was adapted by several authors in order to deal with the formation of German and Dutch verb clusters. A prominent question in the analysis is whether the verbs in the cluster combine with their complements in the same fashion as heads combine with their complements outside the cluster. Hinrichs & Nakazawa (1994) and Bouma & van Noord (1998) treat verb clusters as regular head-complement constructions, whereas Rentier (1994), Kathol (2000), and Müller (2002) introduce a separate valence feature to model verb clusters. While the majority of the analyses presented in this chapter involve binary-branching structures, Bouma & van Noord (1998) provide an analysis based on flat structures.

In this chapter the focus was on presentation, rather than evaluation. In chapter 6, it will be argued that subject raising and complement raising are different phenomena that had better not be dealt with by the same mechanism.

Part II

Corpus Study

CHAPTER 4

Treebank mining

The last decades the use of corpora containing authentic language samples has been of major importance in the study of language.[1] Within theoretical linguistics, corpus examples may provide empirical evidence for the linguistic constructions under investigation, which may support or refute certain assumptions made by the theory. Corpora are not only a source of linguistic constructions; they also allow a quantitative investigation of language, providing insight into the absolute and relative frequencies of the constructions. In the research presented here both of the linguistic applications of corpus data are used. Chapter 5 presents a quantitative investigation of cluster formation in Dutch, while in chapters 6 and 7 corpus examples will be used to support the linguistic analysis and as a means of exemplification.

Before turning to the treebank investigation, the data and the tools to query them will be introduced in this chapter. Section 4.1 discusses the relation between treebanks and linguistics, and refers to previous treebank-supported linguistic research. Section 4.2 presents the treebanks that will be employed for the corpus investigation, while section 4.3 discusses the query languages and the search tools that were used to extract the relevant constructions from the treebanks. Section 4.4 concludes.

[1] Besides their application in (corpus) linguistics, corpora are used in several other research fields. For example, corpora are an important resource within natural language processing (NLP), e.g. for the creation of POS taggers and parsers.

4.1 Corpora, treebanks and linguistics

Corpora come in several types, sizes and formats. For instance, one can differentiate synchronic from diachronic corpora, monolingual from multilingual corpora, plain text corpora from annotated corpora, etc. The kind of corpora that are used within corpus linguistics is largely dependent on the subject of the research and the availability of the corpora. For some languages, no or only plain text corpora are available. Extracting information from raw data is very labour-intensive, especially if one is looking for non-lexical phenomena. The use of annotated data makes such tasks easier, provided that one is familiar with the annotation guidelines, since there is little standardization in the labels and data formats used. The level of data annotation ranges from annotations on the lexical level (e.g. lemmatization and POS tagging) to annotations on the syntactic level (e.g. dependency relations, syntactic categories) and the semantic level (e.g. semantic roles), as well as annotations on the sentence or discourse level (e.g. coreference annotation).

Since the research presented here concerns Dutch syntax, the most obvious choice is to use syntactically annotated corpora, also known as *treebanks*. Typically, the phenomena investigated in syntactic research generalize over word forms and/or word order. Such phenomena are hard to extract from *flat corpora*, i.e. raw or POS-tagged corpora, as they do not contain information beyond the word level. An early overview of the treebanks available for a number of languages is given in Abeillé (2003). The compilation, as well as the use and exploitation of treebanks for linguistic research is discussed in Lüdeling & Kytö (2008), and in Kübler & Zinsmeister (2014).

That treebanks are indeed a valuable resource for linguistics has been shown in several studies on a variety of topics. For instance, van der Beek (2005) made use of several Dutch corpora and treebanks to investigate cleft constructions, the dative alternation and determinerless PPs. Bouma (2004) relies on treebanks evidence for the analysis of PP fronting. Bouma & Spenader (2009) make use of parsed data to investigate the distribution of weak and strong reflexives in Dutch. Van Eynde (2009, 2015) has used the CGN treebank to investigate predicative constructions. Other examples of treebank-supported linguistics can be found in several contributions to the proceedings of the *Treebanks and Linguistic Theories* (TLT) conference.

For Dutch, several treebanks have been constructed over the last decades, such as the Alpino Treebank (van der Beek et al. 2002), LASSY (van Noord et al. 2013), CGN (Oostdijk et al. 2002; Hoekstra et al. 2003), and SoNaR (Oostdijk et al. 2013).

The treebank investigation carried out in chapter 5 is mainly based on the CGN treebank and LASSY Small. The aim of the treebank study is to identify the constructions with verb clusters in the treebanks, to collect information with respect to related phenomena such as word order variation and the IPP effect, and to support the linguistic analysis in chapters 6 and 7. The treebank data not only provide quantitative information on the phenomena under investigation; they also function as an empirical base to verify the theoretical assumptions made in the previous and following chapters.

4.2 Treebanks

For the corpus study, the syntactically annotated version of the CGN Core Corpus (henceforth referred to as the *CGN treebank*) and LASSY Small will be used. Those treebanks for respectively written and spoken Dutch each contain about one million tokens. As the corpora are more or less equal in size, they are well-suited for comparing written to spoken language data.

4.2.1 CGN

The Corpus Gesproken Nederlands (CGN) (Oostdijk et al. 2002) is an annotated corpus of spoken Dutch.[2] It consists of recorded speech which is orthographically transcribed, resulting in a corpus of about ten million words.

The CGN Core Corpus is a representative fragment of the overall corpus containing ca. one million words. That corpus contains POS tags (Van Eynde 2004) as well as syntactic annotations (Hoekstra et al. 2003). That syntactically annotated part will be referred to as the *CGN treebank*.

Two thirds of the corpus data consists of Dutch spoken in the Netherlands, whereas one third of the data comprises Dutch spoken in Flanders, the Dutch speaking part of Belgium. The corpus contains both dialogues and monologues, and is further divided into specific genres. The division into subcorpora allows to investigate stylistic variation (e.g. by comparing spontaneous conversations to news reports), as well as regional variation (by comparing Belgian Dutch to Netherlandic Dutch).

[2] `http://lands.let.ru.nl/cgn`

Contents

Table 4.1 presents the contents of the CGN treebank. The label N is used to refer to the Dutch data, while the label V refers to the Belgian (Flemish) data. The labels A to O refer to the different corpus components that are included in the corpus. The parts A to H contain dialogues, whereas the parts I to O consist of monologues. The dialogues mostly contain spontaneous speech (only component F is more or less prepared), while the monologues mainly consist of prepared speech (only component I is spontaneous). SENTENCES refers to the amount of sentences (or utterances) in each component; WORDS refers to the amount of words (excluding punctuation).

Components	**N**		**V**		**Total**	
	SENTENCES	WORDS	SENTENCES	WORDS	SENTENCES	WORDS
A. Spontaneous conversations ('face-to-face')	50,239	302,828	22,881	147,418	73,120	450,246
B. Interviews with teachers of Dutch	2,484	25,724	4,289	34,158	6,773	59,882
C. Telephone conversations (recorded via a switchboard)	11,649	70,084	3,142	19,984	14,791	90,068
D. Telephone conversations (recorded on MD)	0	0	929	6,309	929	6,309
E. Simulated business negotiations	3,123	25,524	0	0	3,123	25,524
F. Interviews/discussions/debates (broadcast)	6,290	75,167	2,617	25,122	8,907	100,289
G. (Political) discussions/debates/ meetings (non-broadcast)	1,166	25,125	543	9,009	1,709	34,134
H. Lessons recorded in the classroom	3,064	26,004	1,395	10,116	4,459	36,120
I. Live (sports) commentaries (broadcast)	2,251	25,002	1,026	10,147	3,277	35,149
J. News reports (broadcast)	2,259	25,084	536	7,686	2,795	32,770
K. News (broadcast)	1,923	25,353	558	7,306	2,481	32,659
L. Commentaries/columns/reviews (broadcast)	1,857	25,082	601	7,431	2,458	32,513
M. Ceremonious speeches/sermons	444	5,190	107	1,894	551	7,084
N. Lectures/seminars	593	14,921	701	8,159	1,294	23,080
O. Read speech	0	0	3,256	44,144	3,256	44,144
Total	**87,342**	**671,088**	**42,581**	**338,883**	**129,923**	**1,009,971**

Table 4.1: Contents of the CGN treebank

The word and sentence counts in Table 4.1 are based on the CGN Treebank version

2.0.1,[3] converted to the Alpino-XML data format (see appendix B).[4]

Each sentence in the corpus has a unique identifier, e.g. [fva400392__6] for the sentence in (4.1).

(4.1) awel 'k ga ne keer een typisch voorbeeld geven.
well I go a time a typical example give
'well, I'll give a typical example.' [CGN, fva400392__6]

The sentence ID refers to the origin of the fragment (in this case V, for the Belgian part of the corpus; the Netherlandic data are indicated by N), the component (in this case A, for the subcorpus containing spontaneous conversations), the fragment number (400392), and the sentence number (6).[5]

Linguistic annotations

The CGN treebank contains POS tags (Van Eynde 2004) as well as syntactic annotations, assigned according to the guidelines in Hoekstra et al. (2003). The resulting syntactic structures can be represented as tree structures, as in Figure 4.1.
Figure 4.1 shows the most important levels of annotation for the corpus study presented in the following chapters, i.e. the labels containing syntactic and lexical information.

Each word in the tree is assigned a lemma, which generalizes over word forms, and a (Dutch) POS tag, denoting the lexical category of the word (e.g. `adj` for adjective, `ww` for verb). Those lexical tags are only added to the terminal nodes of the tree. Besides the short POS tags presented in the tree structure, the treebank furthermore contains the more finegrained CGN tags (Van Eynde 2004). Those tags not only provide information on the lexical category, but also contain morphosyntactic information, such as case, gender, and number.

The syntactic annotations include information on phrasal categories (e.g. `np` for noun phrases) and dependency relations (e.g. `hd` for syntactic heads). The syntactic categories are added to all non-terminal nodes. The `top` node dominates the syntactic structure and the punctuation.

The dependency labels specify the relation between a node and its mother node. Besides those local dependencies, the treebank also contains information on non-local

[3] `http://tst-centrale.org/nl/producten/corpora/corpus-gesproken-nederlands`
[4] `http://www.let.rug.nl/vannoord/Lassy/alpino_ds.dtd`
[5] In the official release, it is not encoded in the identifier whether the sentence occurs in the Dutch or the Flemish data. This information was added afterwards (based on the metadata in the corpus).

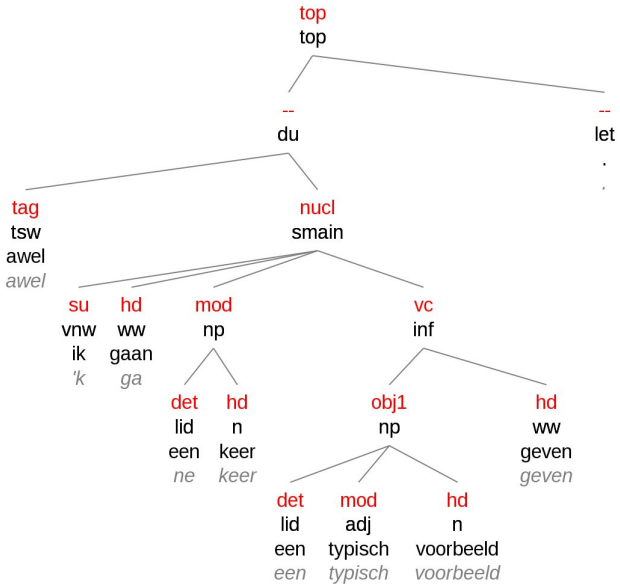


Figure 4.1: Tree representation of a CGN sentence [fva400392__6]

dependencies. For example, in (4.2) the relative pronoun *wat* 'what' is both the head of the relative clause (`rhd`) and the subject of the verb-final clause (`su`).

(4.2) want geef eens een voorbeeld van wat er helemaal niet gebeurd is?
because give once an example of what there completely not happened is
'Could you give an example of what did not happen at all?' [CGN, fnf007268__20]

In the tree structure, such non-local dependencies are indicated by an index (1), as illustrated in Figure 4.2.

Since the trees in the treebank are dependency trees, the order of the terminal nodes in the tree structure does not necessarily correspond to the word order in the sentence. Consider the example in (4.3), which is a case of adposition stranding:

(4.3) kan je daar een voorbeeld van geven?
can you there an example of give
'Can you give an example of that?' [CGN, fvh400225__190]

The corresponding tree structure in Figure 4.3 shows that the pronominal complement *daar* 'there' is a daughter of the PP node, and therefore appears as a sister of the preposition *van* 'of'. The position of the word in the sentence is not indicated in the tree representation, but it is encoded in the underlying XML structure means of the

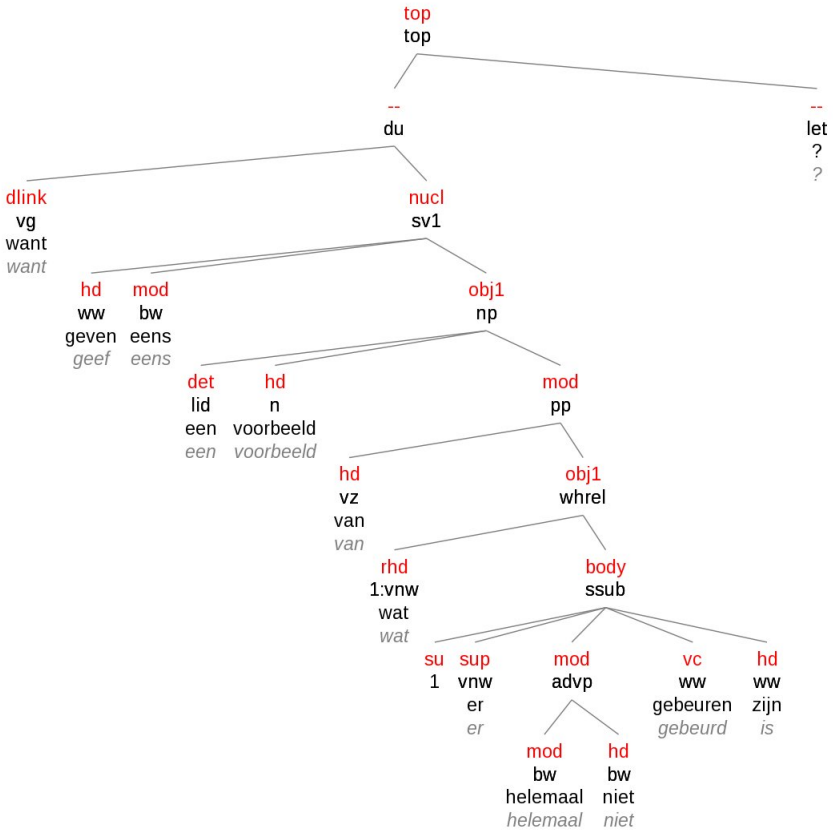


Figure 4.2: Tree representation of a CGN sentence [fnf007268__20]

feature `begin`.[6] For example in (4.3), the `begin` value of *kan* is 0, the value of *je* is 1 etc.

The annotations of the CGN treebank are manually corrected, which makes the treebank a high-quality resource for linguistic research. A validation test showed that the syntactic annotations have an accuracy of 97.53% on sentence level (Fersøe et al. 2006: 39).[7] A detailed overview of the linguistic information included in the CGN treebank is provided in appendix B.1.

[6]For an example of the underlying XML structure, see appendix B.1.

[7]The result is based on the number of sentences in the test set minus the number of sentences containing errors divided by the total number of sentences in the test set.

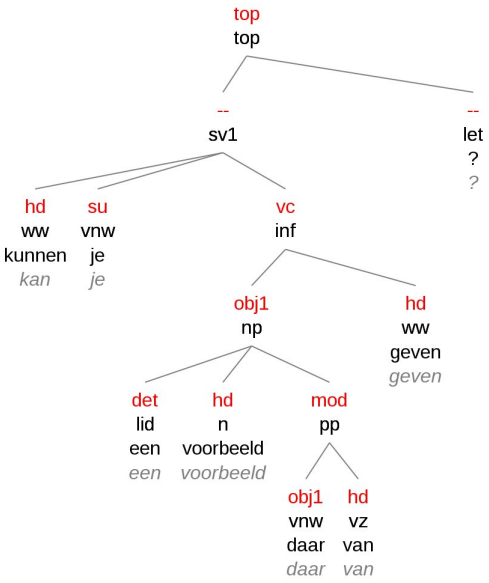


Figure 4.3: Tree representation of a CGN sentence [fvh400225__190]

4.2.2 LASSY

The LASSY treebank (Large Scale Syntactic Annotation of written Dutch) is a syntactically annotated corpus of written Dutch (van Noord et al. 2013).[8] The LASSY project resulted in the construction of two treebanks: LASSY Small and LASSY Large. LASSY Small is a one million word corpus with manually verified annotations, while LASSY Large is a 700 million word corpus with automatically assigned syntactic annotations. For the corpus study presented here LASSY Small is used, since it is complementary to the CGN treebank.

Contents

Table 4.2 provides an overview of the contents of the LASSY Small treebank. The word and sentence counts in the table are based on version 1.1 of the treebank.[9]

Each sentence in the corpus has a unique identifier, e.g. [dpc-med-000678-nl-sen.p.17.s.3] for the sentence in (4.4).

[8] http://www.let.rug.nl/~vannoord/Lassy
[9] http://tst-centrale.org/nl/producten/corpora/lassy-klein-corpus/6-66

COMPONENTS	CONTENTS	SENTENCES	WORDS
DPC	Dutch side of the Dutch Parallel Corpus [dpc][10]	11,716	193,029
Wikipedia	Dutch Wikipedia pages [wiki]	7,341	83,360
WR-P-E	E-magazines [WR-P-E-C], news letters [WR-P-E-E], Teletext pages [WR-P-E-H], Web sites [WR-P-E-I], Wikipedia pages [WR-P-E-J]	14,420	232,631
WR-P-P	Books [WR-P-P-B], brochures [WR-P-P-C], guides and manuals [WR-P-P-E], law texts [WR-P-P-F], newspapers [WR-P-P-G], periodicals and magazines [WR-P-P-H], policy documents [WR-P-P-I], proceedings [WR-P-P-J], reports [WR-P-P-K], surveys [WR-P-P-L]	17,691	281,424
WS-U	auto cues [WS-U-E-A], news scripts [WS-U-T-A], texts for the visually impaired [WS-U-T-B]	14,032	184,611
TOTAL		**65,200**	**975,055**

Table 4.2: Contents of LASSY Small

(4.4) Laat de inkt maar vloeien.
let the ink just flow
'Just let the ink flow.' [LASSY, dpc-med-000678-nl-sen.p.17.s.3]

The sentence ID refers to the subcorpus (in this case DPC-med), the text number (000678), and the location within the text (page 17 sentence 3). The division into subcorpora allows to investigate stylistic variation (e.g. by comparing newspaper articles to law texts).

Linguistic annotations

In contrast to the CGN treebank, LASSY Small was automatically annotated using the Alpino parser (van Noord 2006), a dependency parser for Dutch.[11] In a following step, lemmas and the CGN/D-COI POS tags were automatically assigned using Tadpole (which in the meantime is replaced by Frog), a morpho-syntactic analyzer and tagger for Dutch (Van den Bosch et al. 2007).[12] After the corpus was parsed and tagged, the annotations were manually checked following the LASSY annotation manual (van

[10] Paulussen et al. (2006), `http://www.kuleuven-kulak.be/DPC`
[11] `http://www.let.rug.nl/vannoord/alp/Alpino`
[12] `http://ilk.uvt.nl/frog`

Noord et al. 2011), which is largely based on the CGN guidelines. Finally, the syntactic dependency annotations as well as the POS and lemma annotations were integrated into a single XML representation.

The general lay-out of the treebank is very similar to the CGN treebank, as it contains the same POS tags, and almost the same syntactic annotations. The main annotation difference is the use of indexed nodes in the case of raising and control verbs, as illustrated in Figure 4.4. Since *de inkt* 'the ink' is both the understood object of *laten* 'let' and the subject of the embedded verb *vloeien* 'flow', it is also included as the subject of the verbal complement (VC) in the form of an index node. In CGN, such index nodes are only used for long-distance dependencies, cf. Figure 4.2. A detailed overview of the linguistic information included in LASSY Small is provided in appendix B.2.

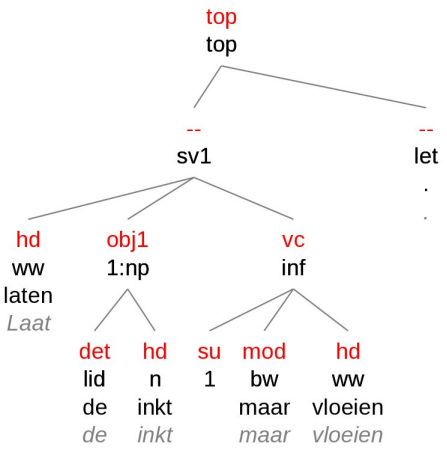


Figure 4.4: Tree representation of a LASSY sentence [dpc-med-000678-nl-sen.p.17.s.3]

Because of the corrections, LASSY Small is a high-quality resource: the sentence-level accuracy is 97.8% (Jongejan et al. 2011: 9–11).

4.3 Querying the treebanks

This section presents the query languages (sections 4.3.1 and 4.3.2) and the search tools that can be used to query the CGN and LASSY treebanks (sections 4.3.3 and 4.3.4).

4.3.1 XPath

There are various query languages that can be employed to extract information from treebanks. The choice usually depends on the data format of the treebanks and on the tools that are used to query them. The data for the treebank study in this thesis were obtained by means of the XPath query language, which is a W3C standard for querying XML documents.[13]

There are several XPath tutorials available. Some general introductions can be found online,[14] but more interesting in this context are the introductions indended for querying treebanks in Alpino-XML format, such as Bouma & Kloosterman (2002) and van Noord et al. (2013). In order to understand the XPath queries used for the treebank investigation, a short introduction will be presented in the remainder of this section.

In each tree of the CGN and LASSY Small treebank, every node in the tree is represented as a `node` element in the XML structure. All the nodes in the treebank can be retrieved using the query in (4.5).

(4.5) `//node`

By adding constraints to those `node` elements, one can look for more specific constructions. For example, the query in (4.6) matches all (lexical) nodes with the POS tag WW (verb).

(4.6) `//node[@pt="ww"]`

Several constraints can be added to the same node. For example, if one is looking for constructions in which a verb is the head of the phrase, the query in (4.7) can be used.

(4.7) `//node[@pt="ww" and @rel="hd"]`

Moreover, one can query for entire tree structures by navigating through the XML structure. The double dots are used to go up one level in the tree. For example, the query in (4.8) matches constructions in which a verb is the head of the phrase, and in which it also has a VC (verbal complement) as a sister node.

(4.8) `//node[@pt="ww" and @rel="hd"]/../node[@rel="vc"]`

[13] `http://www.w3.org/TR/xpath/`

[14] See for example `http://www.w3schools.com/xpath/`.

The same results can be obtained by using embedded structures, as illustrated in (4.9). That notation resembles the labeled bracketing notation which is commonly used in linguistics.

(4.9) `//node[node[@pt="ww" and @rel="hd"] and node[@rel="vc"]]`

XPath is a flexible query language, which has a broader range of applications than querying linguistic data. The downside, however, is that it not only requires time and effort to familiarize oneself with XPath; it is furthermore necessary to have a thorough knowledge of the annotation guidelines of the treebank(s) in order to formulate the queries. As will be shown in section 4.3.3, the search tool GrETEL aims to overcome this problem by providing the option to automatically generate XPath queries.

4.3.2 XQuery

XQuery is a scripting language for XML documents, which is more complex than XPath, but more powerful as well.[15] In fact, XQuery incorporates XPath. For example, the XQuery script in Figure 4.5 can be used to extract the lemmas of the verbs selecting a VC, as well as their frequency in the CGN treebank.[16]

```
let $input := db:open("CGN")/treebank/alpino_ds (: open treebank :)
//node[node[@rel="hd" and @pt="ww"] and node[@rel="vc"]] (: XPath :)

let $verb := $input/node[@rel="hd" and @pt="ww"]  (: verb node :)
let $nl := "&#10;" (: newline character :)

return
for $tag in distinct-values($verb/@lemma)
let $count := count($verb[@lemma eq $tag]) (: count lemmas :)
order by $tag (: order alphabetically :)
return concat($tag,":",$count,$nl) (: return lemma and frequency :)
```

Figure 4.5: Example XQuery script

General introductions to XQuery can be found online;[17] more examples on querying treebanks by means of XQuery are given in Bouma & Kloosterman (2007).

[15] `http://www.w3.org/TR/xquery/`

[16] Comments are written between (: and :).

[17] See for instance `http://www.w3schools.com/xquery/`.

For the treebank investigation XQuery is mainly used to efficiently refine and sort the data that were extracted by means of XPath queries. The XPath queries used for the treebank study will be presented in chapter 5, as they provide information on the constructions that were extracted from the treebanks, but the XQuery scripts will not, as they do not provide any additional information about the relevant constructions.

4.3.3 GrETEL: An online search engine for treebanks

GrETEL (**Gr**eedy **E**xtraction of **T**rees for **E**mpirical **L**inguistics) is a linguistic search engine for treebanks (Augustinus et al. 2012, 2013). It is the result of Nederbooms,[18] a CLARIN[19] project which aimed at the development of user-friendly tools for the exploitation of treebanks by linguists who are not familiar with language technology. An updated version of the tool was created in the GrETEL 2.0 project.

GrETEL is accessible online,[20] which means that users do not have to install any treebanks or specific software (e.g. a parser) locally. The tool has two search modes: *Example-based querying* and *XPath search*. In the remainder of this section a short introduction to GrETEL will be presented. A more detailed manual and some example case studies are provided on the GrETEL website.

Example-based querying

(Descriptive) linguists are often reluctant towards using treebanks, on the one hand because of the limited user-friendliness of the query languages and search tools, and on the other hand because of the lack of standardization in both treebanks and query languages. In order to compensate for those problems, GrETEL provides the example-based querying option, starting from a natural language example instead of a formal search instruction.

Step 1: Provide an example In the example-based search mode, GrETEL does not require any formal input query. As input, the tool takes something linguists are familiar with: natural language. Since linguists tend to start their research from example sentences, the methodology of example-based querying allows users to search for similar constructions as the example they provide. How similar is for the user to decide.

[18] `http://nederbooms.ccl.kuleuven.be`
[19] `http://clarin.eu`
[20] `http://gretel.ccl.kuleuven.be`

For example, if one is looking for IPP constructions, i.e. constructions in which a verb appears as an infinitive if it is selected by an auxiliary of the perfect, one could feed example (4.10) to the system.

(4.10) ...dat hij Marie heeft horen zingen.
...that he Marie have.FIN hear.IPP sing.INF
'...that he has heard Marie sing.

Step 2: The parse tree GrETEL parses the input example with the Alpino parser and returns the input construction as a syntax tree. The search instruction will be based on this parse tree, so if the syntactic analysis of the example is wrong the user is advised to choose another input example.

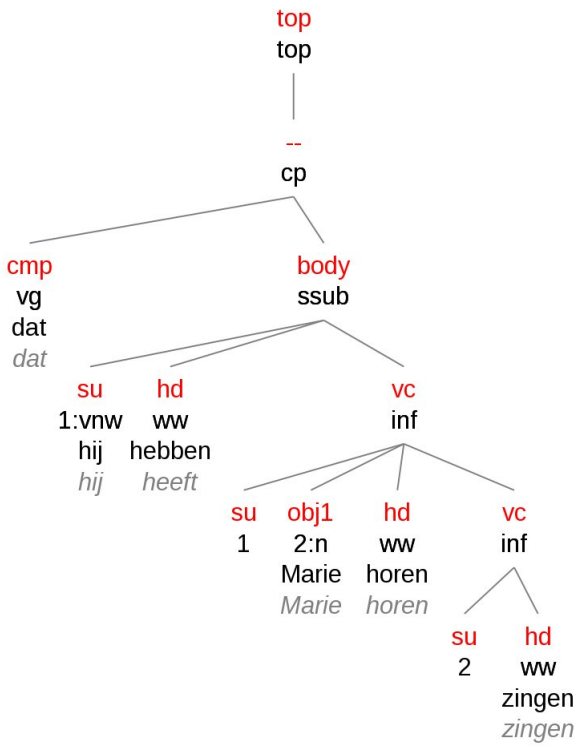


Figure 4.6: Parsed version of the input example

Figure 4.6 shows a correct parse of the input construction in (4.10), so the example can be used for querying the treebanks.

Step 3: The selection matrix In the third step, the user can indicate for each word whether (s)he is interested in the part-of-speech ('word class'), the lemma, or the

(exact) word form. Since we are looking for IPP constructions, the lemma of the perfect auxiliary is relevant, as well as the POS of the infinitives. The other nodes are not relevant for retrieving the target construction. For those words, the 'optional in search' button should be used (see Figure 4.7).

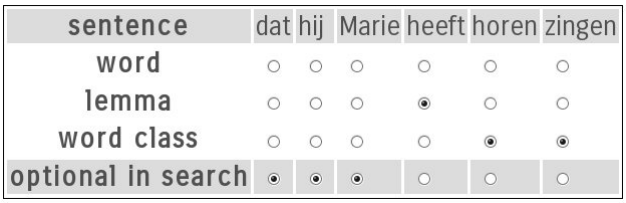


Figure 4.7: Selection matrix

Step 4: Treebank selection The user can choose which treebank (s)he wants to query. Currently one can choose between the CGN treebank, LASSY Small, or SoNaR. For the CGN and LASSY treebanks, it is possible to query the entire treebank, or to query only certain components of a treebank (for instance, only the Belgian parts of CGN). For SoNaR, it is currently only possible to select one component at a time. In this example, the entire CGN treebank is searched.

Step 5: Query overview Based on the information provided in the selection matrix, GrETEL extracts a query tree from the parse tree (Figure 4.8). Besides the lexical information indicated in the selection matrix, the dependency relation (`rel`) and the syntactic category (`cat`) of all relevant nodes is included in the query tree as well.

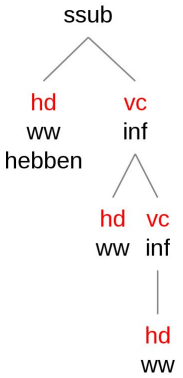


Figure 4.8: Query tree based on the input construction

The query tree is automatically converted into an XPath expression (4.11), which can be used to query the treebank.

(4.11)
```
//node[@cat="ssub" and node[@rel="hd" and @pt="ww" and
@lemma="hebben"] and node[@rel="vc" and @cat="inf" and
node[@rel="hd" and @pt="ww"] and node[@rel="vc" and @cat="inf"
and node[@rel="hd" and @pt="ww"]]]]
```

In the basic search mode, the query is not shown to the user at this stage. In the advanced search mode, users can optionally adapt the XPath query in order to refine or generalize the search instruction.[21]

Step 6: Results In the final step the search results are presented, i.e. the sentences containing the construction at hand. The user can inspect the tree and/or the source XML of the results. It is also possible to download the results in text format. For the query in (4.11), 79 matches in 76 sentences were found in the CGN treebank. Some examples are given in (4.12).

(4.12) a. dus uh ik vermoed dus dat de journalisten de film twee keer hadden
thus uh I assume thus that the journalists the movie two times have.FIN
moeten zien.
must.IPP see.INF
'so uh I assume that the journalists had to watch the movie twice.' [CGN, fvf600243__10]

b. dus 't is niet dat we daar echt mensen hebben leren kennen.
so it is not that we there really people have.FIN learn.IPP know.INF
'so it is not the case that we really have learned to know people over there.' [CGN, fvb400155__296]

Note that the 'Gr' in GrETEL stands for *greedy search*.[22] This means that the matches may include constructions in which nodes appear between the nodes defined in the query tree. An example is the separable verb particle *mee* 'with' in (4.13), which appears in between the verbs of the IPP construction:

[21] Using the XPath expression in (4.11) the results are limited to verb-final constructions (SSUB), and constructions in which the finite verb selects bare infinitives. The query thus does not take into account IPP constructions with *te*-infinitives (TI), or verb-initial constructions (SMAIN and SV1).

[22] The notion *greedy* is used in a similar way as pattern matching with regular expressions, see a.o. Jurafsky & Martin (2009: 56); XPath expressions are greedy in the sense that they match with as much of a tree pattern as they can.

(4.13) ... dat Joegoslavië eigenlijk uh niet had mee mogen doen.
... that Yugoslavia actually uhm not have.FIN with may.IPP do.INF
'... that Yugoslavia in fact was not allowed to join.' [CGN, fnl007393__74]

Even though the primary goal of GrETEL is to attract novice users, advanced users can also benefit from the tool, as it is easier to adapt an XPath query than to build it from scratch.

For instance, if one wants to include verb-initial sentences, constructions with the auxiliary *zijn* 'be', and constructions in which the IPP verb selects a *te*-infinitive or a terminal VC node as well, one could generalise the query in (4.11) to the query in (4.14):

(4.14)
```
//node[@cat and node[@rel="hd" and @pt="ww" and
(@lemma="hebben" or @lemma="zijn")] and node[@rel="vc" and
@cat="inf" and node[@rel="hd" and @pt="ww"] and node[@rel="vc"
and (@cat="inf" or @cat="ti" or @pt="ww")]]]
```

By underspecifying the `@cat` feature for the top node, both verb-initial and verb-final sentences will be included in the results. The `or` operator is used to include both *hebben* 'have' and *zijn* 'be', and to extend the type of VC complement. Terminal VCs have the `@pt="ww"` feature and *te*-infinitives contain the `@cat="ti"` feature.

The query in (4.14) returns 792 matches in 777 sentences in CGN. Some examples are presented in (4.15). The verb-initial sentences clearly show the greedy nature of GrETEL, since in those clauses the finite verb and the infinitives are usually not adjacent. Such constructions would be hard to extract in an efficient way using a flat corpus.

(4.15) a. de organisatie heeft een groot aantal partytenten uit België laten
the organization has a big number party-tents from Belgium let.IPP
komen.
come.INF
'the organization has got a large number of party tents coming from Belgium.' [CGN, fnk005386__3]

b. ik heb m'n nicht proberen te bellen want die uh ...
I have my cousin try.IPP to call.INF because that one uh
'I have tried to call my cousin because she erm ...' [CGN, fna000628__73]

If one does not want to tinker with the XPath query, one can also build separate queries using slightly different input examples.

XPath Search

Besides querying treebanks by example, it is also possible to query the treebanks by means of an XPath expression straightaway, using the *XPath Search* version of GrETEL. Similar to Dact (see section 4.3.4), the user has to provide the XPath queries. One could use for instance the queries presented in section 4.3.1, but one can also generate a query using the example-based method and adapt it in the XPath search mode.

4.3.4 Stand-alone search tools

Besides GrETEL there are a number of treebank search tools that work offline. In contrast to an online tool like GrETEL, they need to be locally installed on a computer, together with the treebanks.

TIGERSearch

TIGERSearch (Lezius 2002) is a graphical user interface for querying treebanks encoded in the TIGER-XML data format (see appendix B for an example XML file). The data can be queried using the TIGER language.[23] Although the original format of the CGN treebank is TIGER-XML, TIGERSearch was not used for querying CGN, on the one hand because the software is outdated (i.e. it is no longer supported on current operating systems), and on the other hand because the LASSY treebank cannot be queried using TIGERSearch. Since it is possible to query both (the converted version of) the CGN treebank and LASSY Small with XPath, that query language was chosen to extract the data from the treebanks.

Dact

Dact (van Noord et al. 2013) is a graphical user interface to query treebanks encoded in the Alpino-XML data format (see appendix B for an example XML file).[24] It can be used to query both the LASSY treebank and the converted version of the CGN treebank.[25]

[23] For a manual on both TIGERSearch and the TIGER query language, see `http://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/TIGERSearch/manual_html.html`.

[24] `http://rug-compling.github.com/dact`

[25] Dact has largely the same functionality as dtsearch (Kloosterman 2007), which is an (older) command-line interface for querying XML treebanks.

Treebanks included in Dact can be queried with XPath. In contrast to GrETEL, it does not come with a query building functionality. It is possible, however, to use the queries created with GrETEL. If one enters an XPath query in Dact, it is possible to inspect the matching tree structures, in which the nodes that match the XPath query are highlighted.

Besides browsing the tree structures, it is furthermore possible to get frequency information on the constructions that match the query. Moreover, if the query matches any lexical items (i.e. terminal nodes), it is possible to extract a list of the matching lemmas, as well as their frequency.

BaseX

BaseX is a native XML database engine.[26] It comes with a GUI, but in contrast to the other tools presented in this section, BaseX does not specifically aim at querying (linguistic) treebanks, so it has no module to visualize syntax trees.

Similar to the tools described earlier in this section, BaseX can be used for querying XML trees by means of XPath. In addition, it also supports XQuery, which is more flexible in comparison to XPath (cf. section 4.3.2).

4.4 Conclusion

This chapter pointed out the use of corpora, and more specifically treebanks for linguistic research. The treebanks used for the corpus investigation in chapter 5 were presented, and an overview was given of the different query languages and search tools for Dutch treebanks. GrETEL was mainly used in order to create the XPath queries and for finding relevant constructions, whereas Dact and BaseX were used for refining and sorting the results. Those search tools all make use of the XPath query language.

[26] `http://basex.org/`

CHAPTER 5

A treebank-supported investigation of verb clusters

In this chapter it will be shown how verb clusters can be retrieved in the CGN and LASSY Small treebanks.[1] The aim of this chapter is to provide an overview of the attested clusters and to identify the clustering verbs in order to define which constructions the analysis proposed in the following chapters should account for. In addition, frequency information is collected, which allows to differentiate typical verb clusters and clustering verbs from less common clusters and clustering verbs. The treebank observations will be compared to previous research that was discussed in chapter 1.

Even though the treebank annotations do not contain a separate tag for clustering verbs, it is possible to automatically extract clustering constructions using the relevant queries. As will be shown in section 5.1, the clustering verbs are a subset of the verbs selecting a verbal complement (VC). Once this maximal set of potentially clustering constructions is extracted, the data will be narrowed down to the constructions containing verb clusters. Section 5.2 discusses verb clusters with bare infinitives and/or a past participle. After extracting the relevant constructions, the treebank observations about the frequency, length, and word order will be presented. Furthermore, the set of clustering verbs, i.e. the verbs that trigger cluster formation, will be identified and classified. Section 5.3 discusses constructions with a *te*-infinitive. They are dealt with in a separate section, as clusters containing a *te*-infinitive are harder to differentiate from non-clustering constructions, as was argued in section 1.10. Section 5.4 dis-

[1] In examples, figures and tables presented in this and the following chapters the shorthand notation 'CGN' and 'LASSY' will be used to refer to the CGN and LASSY Small treebanks (rather than the entire CGN corpus or the LASSY Large treebank).

cusses the relation between the IPP effect and cluster formation, and describes which verbs occur as (optional or obligatory) IPP verbs in the data. Section 5.5 identifies constructions containing cluster creepers, and discusses which types of non-verbal elements occur as cluster creepers. Section 5.6 concludes and gives an overview of the clustering verbs encountered in the data, showing the distribution with respect to the type of verbal complement they select and their occurrence in IPP constructions.

5.1 Constructions with a verbal complement

The treebanks do not contain a specific label for clustering verbs (nor for their verbal complements), but there is a tag for verbal complements (VC) in general. Figure 5.1 presents the Alpino parses of the sentences in (5.1), showing that both verbal complements in the cluster and verbal complements in the Nachfeld receive the tag VC.

(5.1) a. ... dat hij het nijlpaard heeft proberen te vangen.
... that he the hippo has try.IPP to catch.INF
'... that he has tried to catch the hippo.'

b. ... dat hij heeft geprobeerd het nijlpaard te vangen.
... that he has try.PSP the hippo to catch.INF
'... that he has tried to catch the hippo.'

Since all clustering verbs select a verbal complement, the set of VCs includes the set of verbal complements selected by clustering verbs. The VCs can be retrieved by means of the query in (5.2a). The query in (5.2b) returns the VC nodes which are selected by a head sister.

(5.2) a. `//node[@rel="vc"]`

b. `//node[node[@rel="hd"] and node[@rel="vc"]]`

The CGN treebank contains 39,233 VCs, of which 39,006 have a head sister (i.e. the VC selector).[2] LASSY Small contains 43,119 VCs, which all have a head sister. The VCs in CGN occur in 29,214 out of the 129,923 utterances (22.49%), while in LASSY they occur in 28,482 out of the 65,200 sentences (43.68%).

[2]Most of the VCs without head sister occur in elliptic or incomplete utterances.

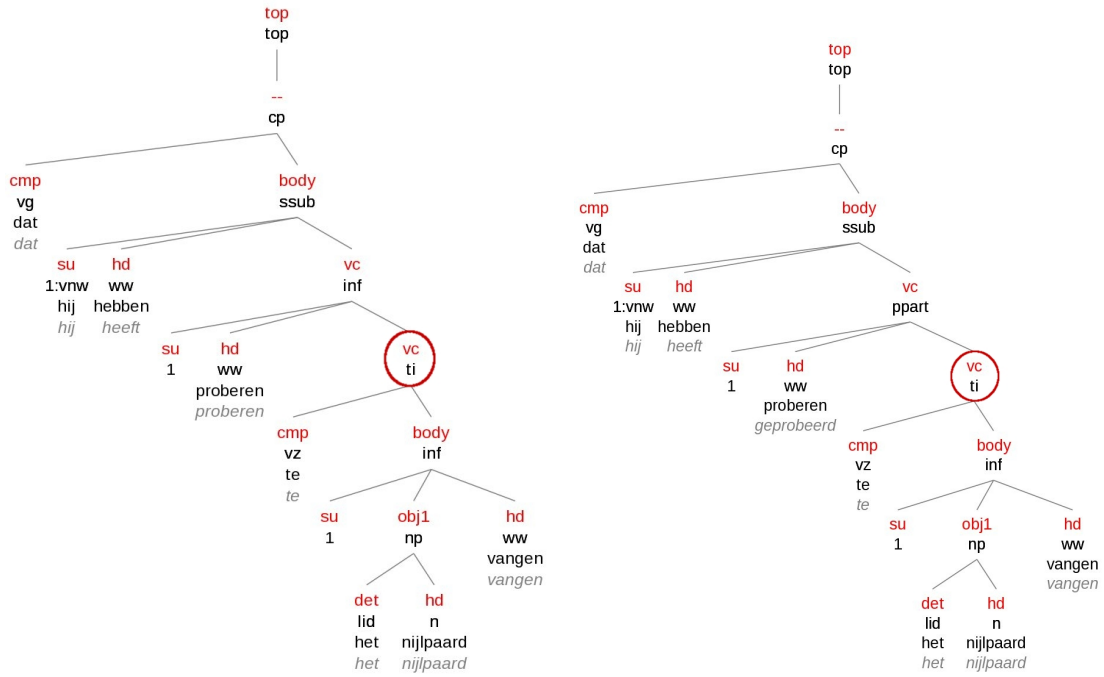


Figure 5.1: Representation of verbal complements (VC)

Only verbal VC selectors are candidates to trigger verb clusters.[3] They can be found by adding the constraint `@pt="ww"` to the head node in (5.2b), resulting in query (5.3).

(5.3) `//node[node[@rel="hd" and @pt="ww"] and node[@rel="vc"]]`

In CGN, the verbal VC selectors account for 98.2% of all VC selectors, while in LASSY they account for 92.3% of the VC selectors.[4] The figures are summarized in Table 5.1. The constructions containing verb clusters are a subset of the VCs selected by a verbal head. The non-clustering constructions, such as constructions in which the VC is a complementizer phrase, can be eliminated from the data set by using more specific queries. In the following sections, the data set will be narrowed down in order to identify the constructions with verb clusters.

[3]Non-verbal VC selectors mainly include nouns, adjectives, adverbs and prepositions selecting a *te*-infinitive or a complementizer phrase, e.g. *het ogenblik dat de satelliet voorbijkomt* 'the moment the satellite comes past', *blij haar te zien* 'happy to see her'.

[4]The query in (5.3) does not return all verbal selectors in the treebanks. Besides the selectors with a WW tag, selectors that are part of a conjunction (tagged CONJ) may be verbal, as well as some of the selectors tagged SPEC in CGN. The latter group mainly consists of interrupted or unintelligible speech. As those selectors are a minority of the VC selectors, they are not taken into account.

	CGN		LASSY		**Sum**	
	VCS	%	VCS	%	VCS	%
VERBAL SELECTOR	38,302	98.2	39,808	92.3	78,110	95.1
NON-VERBAL SELECTOR	704	1.8	3311	7.7	4,015	4.9
Sum	**39,006**	**100**	**43,119**	**100**	**82,125**	**100**

Table 5.1: VC selection in CGN and LASSY Small

5.2 Verb clusters with bare infinitives and/or a past participle

In chapter 1 a verb cluster was defined as a sequence of two or more verbs occurring in the second pole. This section discusses the treebank observations with respect to verb clusters without a *te*-infinitive, i.e. clusters with a finite verb, one (or more) bare infinitive(s) and/or a past participle. It was motivated that verbs selecting a *te*-infinitive can be clustering, non-clustering or ambiguous (cf. section 1.10). Therefore, constructions containing *te*-infinitives will be dealt with in a separate section.

Clusters containing a past participle will be referred to as *participial clusters*, while clusters without a participle will be referred to as *infinitival clusters*.[5] Section 5.2.1 discusses how the relevant constructions were extracted from the treebanks. Section 5.2.2 presents the treebank results with respect to word order variation in participial and infinitival verb clusters. Section 5.2.3 provides an overview and discussion of the clustering verbs occurring in participial and infinitival verb clusters. Section 5.2.4 concludes.

5.2.1 Extracting the constructions

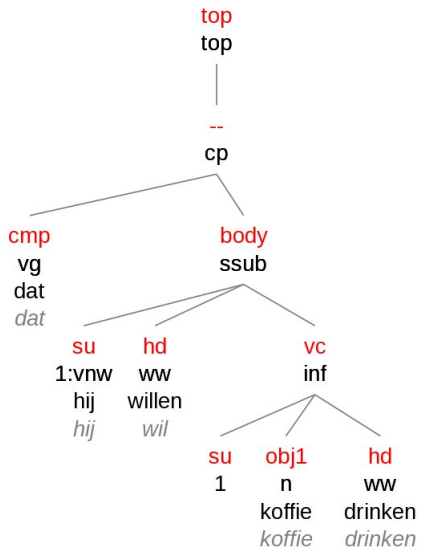
For the extraction of participial and bare infinitival clusters from the treebanks, XPath queries were generated using the example-based method in GrETEL. For example, the input construction in (5.4) was used to automatically derive the query in (5.5a).[6] (5.5b) is a visual representation of the query, i.e. a subtree of the parse in (5.4).

[5] Participial clusters can contain a finite verb and/or bare infinitives as well, while infinitival clusters can only contain a finite verb in addition to the bare infinitive(s).

[6] In order to derive the XPath query, `word class` was indicated for the verbs in the selection matrix in GrETEL.

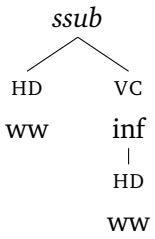
(5.4) a. ... dat hij koffie wil drinken.
... that he coffee wants drink.INF
'... that he wants to drink coffee.'

b.



(5.5) a.
```
//node[@cat="ssub" and
      node[@rel="hd" and @pt="ww"] and
      node[@rel="vc" and @cat="inf" and
         node[@rel="hd" and @pt="ww"] ] ]
```

b.

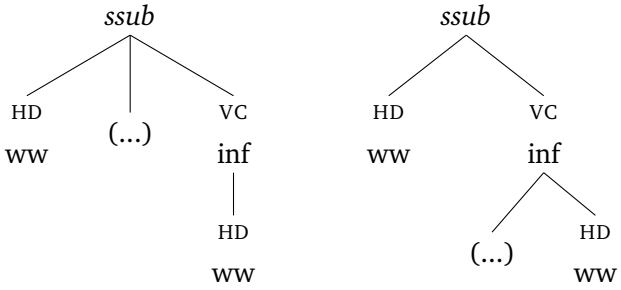


The query in (5.5a) extracts verb-final constructions (`ssub`) with a verb (`ww`) as head daughter (HD) and a verbal complement (VC) in the form of a bare infinitive phrase (`inf`). The XPath engine does not take into account the order of the nodes; for the query in (5.5a) it also returns constructions in which the selecting verb follows the infinitive (e.g *drinken wil* 'drink wants').

Furthermore, the XPath engine performs a *greedy* search, i.e. queries like (5.5a) not only return constructions in which the two verbs are adjacent, but also the constructions in which another element intervenes between the two verb forms, i.e. con-

structions with cluster creepers such as *wil koffie drinken* 'wants coffee drink'. So constructions like the ones in (5.6) are included as well.[7]

(5.6)



The XPath expressions can be further specified or generalized by adding or removing constraints. For example, by adding the constraint `@wvorm="pv"` (for *persoonsvorm* 'finite verb') the query will return constructions in which the selecting verb is finite. Greedy search furthermore means that the query in (5.5a) returns all matches containing *at least* a verb and a bare infinitive, so it will also return constructions with more than two verb forms. In order to compare for instance two-verb clusters to three-verb clusters (cf. infra), one needs to control the cluster length. This can be done by adding a constraint stating that the VC node should have no other VC daughter, using the `not()`-operator. The resulting query is shown in (5.7).

(5.7)
```
//node[@cat="ssub" and
        node[@rel="hd" and @pt="ww" and @wvorm="pv"] and
        node[@rel="vc" and @cat="inf" and
           node[@rel="hd" and @pt="ww"] and
           not(node[@rel="vc" and (@cat="inf" or @cat="ti"
           or @cat="ppart" or @pt="ww")]) ] ]
```

The queries in (5.5a) and (5.7) look for non-terminal VC nodes, i.e. the VC nodes containing more than one daughter. Since in the LASSY treebank index nodes are used to indicate raising and control constructions, almost all VC nodes are non-terminal, cf. the tree in (5.4) and Figure 4.4 in section 4.2.2. In the CGN treebank VCs are non-terminal if the VC node consists of more than one word, e.g. a verb and a direct object, such as *koffie drinken* 'drink coffee' in (5.4). If the VC node consists of one word, it is represented as a terminal node in the treebank, cf. the tree in Figure 4.2 in section 4.2.1. To retrieve the constructions with terminal VCs, the query in (5.8a) is used.[8]

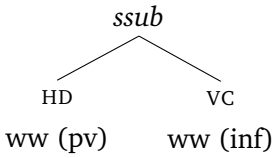
[7] The label (...) stands for any sequence of nodes that may occur in that position.

[8] The `not()` condition need not be stated here, since a terminal (VC) node cannot have any (VC) daughters.

The query tree is presented in (5.8b).

(5.8) a.
```
//node[@cat="ssub" and
      node[@rel="hd" and @pt="ww" and @wvorm="pv"] and
      node[@rel="vc" and @pt="ww" and @wvorm="inf"] ]
```

b.

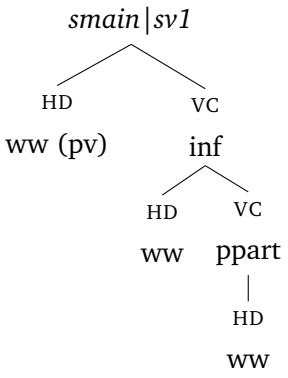


The other participial and bare infinitival clusters can be found by means of adaptations and extensions of the queries presented in (5.7) and (5.8a). Constructions with more than two verb forms in the cluster have one or more VC nodes embedded under the VC. Verb-initial constructions can be retrieved by changing the label `ssub` to `smain` for verb-second clauses or to `sv1` for verb-first clauses. The label `ppart` is used for non-terminal past participles (past participial phrases), whereas the Dutch label `vd` (for *voltooid deelwoord* 'past participle') is used for terminal nodes.

For example, the query in (5.9a) returns V-initial constructions with a finite verb, a bare infinitive and a past participle.[9]

(5.9) a.
```
//node[(@cat="smain" or @cat="sv1") and
      node[@rel="hd" and @pt="ww" and @wvorm="pv"] and
      node[@rel="vc" and @cat="inf" and
         node[@rel="hd" and @pt="ww"] and
         node[@rel="vc" and @cat="ppart" and
            node[@rel="hd" and @pt="ww"] ] ] ]
```

b.



[9] Also in this case the `not()`-function need not be stated. As the past participle is the last element in the order of selection of the cluster, it does not matter whether it has any embedded (extraposed) VC nodes.

5.2.2 Cluster types and word order variation

The extraction of the bare infinitival and participial clusters resulted in 21,029 constructions, i.e. 9,845 in the CGN treebank and 11,184 in LASSY Small. Those constructions are a subset (26.92%) of the 78,110 constructions in which a verb selects a VC (see Table 5.1 in section 5.1). In LASSY, the majority of the constructions contains a past participle (64.6%), whereas in CGN, the infinitival clusters occur more frequently (58.2%), see Table 5.2. For each cluster type, the number of occurrences is the sum of the queries for non-terminal and terminal VCs. The classification of the data presented in this section was done automatically. Only constructions with a low frequency and a non-canonical word order were manually checked and reclassified if necessary.[10]

	CGN		LASSY		**Sum**	
	CLUSTERS	%	CLUSTERS	%	CLUSTERS	%
PARTICIPIAL	4,113	41.8	7,224	64.6	11,337	53.9
INFINITIVAL	5,732	58.2	3,960	35.4	9,692	46.1
Sum	**9,845**	**100**	**11,184**	**100**	**21,029**	**100**

Table 5.2: Participial and infinitival clusters in CGN and LASSY Small

Table 5.3 presents an overview of the clusters according to their length, showing a similar division in both the CGN and LASSY Small treebank. As expected, the majority of the clusters consist of only two verbs (more than 90% of the clusters in both treebanks). The clusters with three verbs account for a bit less than 10% of the constructions, the clusters with four verbs account for less than 0.5% of the participial and infinitival clusters. Neither of the treebanks contains clusters with more than four verbs. Although such constructions are well-formed, the data indicate that they are avoided in actual language use.

Tables 5.4 till 5.6 present the results per cluster length, including the word order variation within the clusters.[11] The order of selection and the word form of the verbs

[10] For instance, two constructions were classified as FINITE_1 INF_3 INF_2, but in one construction a past participle is erroneously tagged as an infinitive (i.e. *zouden gedaan hebben* 'would done have' in CGN, fvb400118__395).

[11] Since it is hard to determine the linear order of the cluster in an elegant way using XPath, XQuery was used to determine the word order.

	CGN		LASSY		**Sum**	
	CLUSTERS	%	CLUSTERS	%	CLUSTERS	%
2 VERBS	9,003	91.4	10,071	90.1	19,074	90.7
3 VERBS	806	8.2	1,078	9.6	1,884	9.0
4 VERBS	36	0.4	35	0.3	71	0.3
Sum	**9,845**	**100**	**11,184**	**100**	**21,029**	**100**

Table 5.3: Length of the participial and infinitival clusters

in the cluster is indicated in the rows, while the linear order is given in the columns. The constructions in which the finite verb is included in the cluster are verb-final, e.g. combinations of a finite verb (FIN) and a past participle (PSP), whereas constructions in which the finite verb is not included in the cluster are verb-initial, e.g. combinations of a bare infinitive (INF) and a past participle.

The aim of this overview is to present the word order variation within the clusters as observed in the treebanks, in order to identify which constructions should be accounted for in the syntactic analysis. The factors that influence word order variation will not be taken into account in this research. They are discussed in amongst others De Sutter (2005, 2009), Coussé (2008), and Bloem et al. (2014).

In section 5.2.1 it was mentioned that the treebank results also contain instances of cluster creeping. In the remainder of this section, only the position of the verbs with respect to the other verbs in the cluster will be considered. The interruption of the clusters by non-verbal material will be discussed in section 5.5.

	CGN		LASSY		**Sum**
	1-2	2-1	1-2	2-1	
FINITE_1 PSP_2	1,664	1,626	3,544	1,519	8,353
INF_1 PSP_2	127	355	830	530	1,842
FINITE_1 INF_2	3,472	43	2,989	6	6,510
INF_1 INF_2	1,714	2	653	0	2,369
Sum	**6,977**	**2,026**	**8,016**	**2,055**	**19,074**

Table 5.4: Two-verb clusters in CGN and LASSY Small

The figures in Table 5.4 indicate that the past participle may occur as the first or

second element in the cluster, irrespective of whether the verb selecting the participle is a finite verb or a bare infinitive. The infinitival clusters, on the contrary, show a clear preference for the canonical (ascending) order. Some exceptions are given in (5.10).[12] The two INF_2 INF_1 examples in CGN occur in the same context and even contain the same verbs, i.e. *surfen leren* 'learn to surf'. One example is given in (5.10c).

(5.10) a. . . . een apart verkeersreglement voor Vlaanderen en Wallonië dat is wat
. . . a separate traffic-regulations for Flanders and Wallonia that is what
ons redden$_2$ gaat$_1$.
us save will
'. . . separate traffic regulations for Flanders and Wallonia is what will save us.' [CGN, fvl600253__22]

b. . . . om ervoor te zorgen dat dit nooit meer gebeuren$_2$ zal$_1$.
. . . to there-for to make-sure that this never again happen will
'. . . in order to make sure that this will never happen again.' [LASSY, dpc-vhs-000725-nl-sen.p.7.s.4]

c. dan kan je toch ook niet goed surfen$_2$ leren$_1$ als je dat niet lukt?
then can you actually also not good surf learn if you that not succeed
'if you don't succeed in that you actually can't learn to surf properly?' [CGN, fna000836__57]'

Table 5.5 presents an overview of three-verb participial and infinitival clusters. Out of the six logical orders (3!), four orders occur in the data. As expected, the 2-1-3 order does not occur in the data, as it is reported to be ungrammatical in all Dutch dialects, see for example Barbiers (2005) and Wurmbrand (2005). The strictly descending 3-2-1 order, typical for the dialects spoken in the northern provinces of the Netherlands, is also absent in the treebank data.

Similar to the clusters consisting of two verbs, participial three-verb clusters show more variation compared to infinitival three-verb clusters. The figures in Table 5.5 show that the past participle can occur at any position in the cluster, but the order of the other verbs in the cluster is always ascending. Some examples are given in (5.11).

[12] As some treebanks instances are very long, several examples are shortened for the discussion in this and the following chapters, which is indicated by '. . . '. The complete examples can be retrieved from the treebanks by means of the sentence IDs.

	CGN				LASSY				**Sum**
	1-2-3	3-1-2	1-3-2	2-3-1	1-2-3	3-1-2	1-3-2	2-3-1	
FINITE_1 INF_2 PSP_3	89	139	63	0	395	262	48	0	996
INF_1 INF_2 PSP_3	4	18	6	0	33	29	4	0	94
FINITE_1 INF_2 INF_3	430	0	1	7	298	0	0	0	736
INF_1 INF_2 INF_3	48	0	0	1	9	0	0	0	58
Sum	**571**	**157**	**70**	**8**	**735**	**291**	**52**	**0**	**1,884**

Table 5.5: Three-verb clusters in CGN and LASSY Small

(5.11) a. In de tussentijd zouden de twee belangrijkste getuigen in
in the meantime would the two most-important witnesses in
Nederland moeten$_1$ worden$_2$ gehoord$_3$.
the-Netherlands must be heard
'In the mean time the two most important witnesses in the Netherlands would have to be heard.' [LASSY, WS-U-E-A-0000000008.p.36.s.3]

b. er zijn toch zo'n paar boeken die ge moet$_1$ gelezen$_3$ hebben$_2$ in
there are actually such couple books that you must read have in
uw leven.
your life
'actually there are a couple of books that you should have read in your life.' [CGN, fva400503__11]

c. Diversiteit in onze samenleving zou nog veel meer benadrukt$_3$ moeten$_1$
diversity in our society should still much more focussed must
worden$_2$.
be
'Diversity in our society should be much more focussed on.' [LASSY, WR-P-P-G-0000000019.p.5.s.14]

With respect to the infinitival clusters the spoken data show some variation, as illustrated in (5.12), but in the written data they exclusively appear in the ascending order.

(5.12) a. ...want ik zou op dat moment geen uh lesvoorbereidingen in uh
...because I would on that moment no uh course-preparations in uh
kunnen$_2$ maken$_3$ hebben$_1$ met de PC hè.
can make have with the PC hè
'...because I wouldn't have been able to make course preparations with the computer at that moment.' [CGN, fvb400165__167]

b. . . . terwijl dat 'k ik naar buiten gaan$_2$ kijken$_3$ ben$_1$.
. . . while that I I to outside go look am
'. . . while I was going to look outside.' [CGN, fva400388__48]

c. we zijn ook met een organisatie uh waar vreemdelingen mogen$_1$
we are also with an organisation uh where foreigners are-allowed
logeren$_3$ komen$_2$ thuis.
stay-over come home
'we are also in an organisation uh in which foreigners are allowed to stay over.'
[CGN, fvb400117__451]

All examples in (5.12) are clearly instances of (spontaneous) spoken Dutch. Haeseryn et al. (1997) label infinitival 2-3-1 constructions as regional Dutch. The eight treebank examples all occur in the Belgian data. The construction is only possible if the hierarchically highest verb is an instance of *hebben* 'have' or *zijn* 'be', i.e. the IPP constructions (Haeseryn et al. 1997: 1066), which is indeed the case in the treebank examples. The 1-3-2 construction is even more remarkable, as it is the only infinitival construction in which the second and the third verb are in descending order. According to Barbiers (2005), this word order can also occur in IPP constructions but no such instances were found in the treebanks. Furthermore, the data do not contain infinitival 3-1-2 constructions, which are also encountered in constructions with two modal verbs in Barbiers (2005).

Compared to other studies on word order variation in three-verb clusters, such as Wurmbrand (2004) and Barbiers (2005, 2008), the treebank data show relatively little word order variation, especially in the case of infinitival clusters. This can be attributed to the fact that the treebanks do not contain dialect material, the variation shown is only regional and stylistic variation within Standard Dutch.

Clusters with four verbs are scarce in comparison to two- and three-verb clusters. Only finite four-verb clusters were found, as shown in Table 5.6. Out of the 24 logical orders (4!), only four are attested.

	CGN				LASSY				**Sum**
	1-2-3-4	4-1-2-3	1-2-4-3	1-3-2-4	1-2-3-4	4-1-2-3	1-2-4-3	1-3-2-4	
FINITE$_1$ INF$_2$ INF$_3$ PSP$_4$	4	11	6	1	18	10	2	0	52
FINITE$_1$ INF$_2$ INF$_3$ INF$_4$	13	0	0	1	5	0	0	0	19
Sum	**17**	**11**	**6**	**2**	**23**	**10**	**2**	**0**	**71**

Table 5.6: Four-verb clusters in CGN and LASSY Small

In LASSY Small, all finite and infinitival verb forms occur in ascending order, while the position of the participle is variable, as shown in (5.13). According to Haeseryn et al. (1997: 1071), participles can occur at any position in four-verb clusters, but no instances of participial 1-4-2-3 clusters were found in the treebank data. This may be due to data sparseness.

(5.13) a. . . . weg van om het even welk gevaar dat door X-6 zou_1 $kunnen_2$
. . . away from to the even which danger that by X-6 would can
$achtergelaten_4$ $zijn_3$.
left-behind be
'. . . away from any danger that by X-6 could have been left behind.' [LASSY, WR-P-E-I-0000051419.p.35.s.5]

b. Het communistische en het democratische blok verschilden scherp van
The communist and the democratic bloc differed sharply of
inzicht over de vraag hoe een staat $ingericht_4$ zou_1 $moeten_2$ $worden_3$.
insight about the question how a state organised should must be
'The communist and the democratic bloc sharply disagreed on the question how a state should be organised.' [LASSY, WR-P-E-I-0000039081.p.1.s.12]

The CGN treebank contains two constructions in which the infinitives do not show the ascending order:

(5.14) a. . . . dat uh delen van het lopend onderzoek door de personen die erbij
. . . that uh parts of the ongoing research by the persons that there-with
betrokken waren $moeten_1$ $worden_3$ $kunnen_2$ $voortgezet_4$.
involved were must be can pursued
'. . . that it should be possible to pursue parts of the ongoing research by the people who were involved.' [CGN, fvf600373__85]

b. dit is de pen waar 'k k uhm mee heb_1 $leren_3$ $moeten_2$ $schrijven_4$.
this is the pen where I I uh with have learn must write
'This is the pen with which I had to learn to write.' [CGN, fnj007333__163]

Due to their low frequency, it is hard to draw conclusions about the distribution of four-verb clusters. Still, the data suggest similar generalizations as were observed with respect to three-verb clusters, i.e. bare infinitives have a strong tendency to appear in the hierarchical (ascending) order, while participles show more variation regarding the position they obtain in the cluster.

In general, the treebank observations with respect to word order are in line with the generalizations formulated by Broekhuis & Corver (2015), cf. section 1.7. They

indicate that a grammar of Dutch should at least account for the fact that infinitives typically occur in the ascending order, while the past participles may obtain other positions in the cluster. In addition, a model for spoken/colloquial Dutch should also account for the fact that the finite verb may obtain a cluster-final position in IPP constructions and in two-verb infinitival clusters.

5.2.3 Clustering verbs

Besides the identification of the cluster types and the word order variation within the cluster, it is possible to extract a list of the VC selectors, i.e. the verbs that trigger cluster formation. Table 5.7 lists the verbs selecting a bare infinitival or a participial VC in the treebanks.[13] It presents the selecting verbs and their occurrence as a selector of a bare infinitive (INF) and/or a past participle (PSP). The figures are the sum of the instances found in the CGN treebank and LASSY Small. The total number of constructions is 23,055, which is the sum of $1 \times 19,074$ (two-verb clusters) + $2 \times 1,884$ (three-verb clusters) + 3×71 (four-verb clusters). In the constructions with three- and four-verb clusters, no selecting verbs were encountered that did not appear in two-verb clusters. By classifying those constructions it turned out that in 63 constructions a predicative complement is selected, rather than a verbal complement. They will be discussed at the end of this section. Verbs which cannot select a past participle (resp. bare infinitive) have '–' in the PSP (resp. INF) column.

The verbs listed in Table 5.7 are divided into two broad categories: subject-oriented and object-oriented verbs. The set of subject-oriented verbs contains subject-raising and the subject-control verbs, i.e. verbs which identify their implied subject with the subject of their verbal complement. Object-oriented verbs identify their understood object with the subject of their verbal complement. A formal definition was given in chapter 3 (section 3.2.4). Some verbs appear twice in Table 5.7. *Leren*, for instance, is a subject-oriented verb if it has the meaning 'learn', but object-oriented if it has the meaning 'teach'.

[13] If a verb selects an infinitival VC, this infinitive may be the main verb (e.g. *zingen* 'sing' in *dat hij leert zingen* 'that he learns to sing') or another clustering verb (e.g. *leren* in *hij zal leren zingen* 'he will learn to sing'). If a verb selects a participle, the participle is the main verb.

	LEMMA	INF	PSP	SUM
SUBJECT-ORIENTED VERBS	*worden* 'be'	–	5,190	5,190
	hebben 'have'	177	3,002	3,179
	zijn 'be'	92	2,968	3,060
	kunnen 'can, be able to'	3,030	–	3,030
	moeten 'must, have to'	2,232	–	2,232
	zullen 'will'	2,035	–	2,035
	mogen 'may, be allowed to'	365	–	365
	hoeven 'need'	9	–	9
	dienen 'ought'	–	1	1
	gaan 'go, will'	1,470	–	1,470
	blijven 'remain, continue'	216	–	216
	komen 'come'	154	1	155
	zitten 'sit'	66	13	79
	beginnen 'start, begin'	25	–	25
	wezen 'be in the process of'	15	–	15
	staan 'stand'	26	25	51
	lopen 'walk'	7	–	7
	liggen 'lie'	1	7	8
	willen 'want'	946	–	946
	leren 'learn'	53	–	53
	durven 'dare'	22	–	22
	proberen 'try'	4	–	4
OBJECT-ORIENTED VERBS	*laten* 'let'	536	–	536
	doen 'do'	70	–	70
	zien 'see'	93	5	98
	horen 'hear'	28	–	28
	voelen 'feel'	2	–	2
	helpen 'help'	10	–	10
	leren 'teach'	2	–	2
	hebben 'have'	51	–	51
	krijgen 'get'	–	38	38
	vinden 'think, find'	5	–	5
	Sum	**11,742**	**11,250**	**22,992**

Table 5.7: Verbs selecting a participal and/or infinitival VC

Subject-oriented clustering verbs

The set of subject-oriented verbs consists of the perfect and passive auxiliaries, modal verbs, aspectual verbs, and some subject control verbs. The majority of the subject-oriented verbs listed in Table 5.7 obligatorily select an infinitive as their verbal complement, but some verbs can also select a participial complement.

If they select a main verb, the **perfect and passive auxiliaries** *hebben* 'have', *zijn* 'be', and *worden* 'be' always select a past participle. The fact that those auxiliaries can occur before and after the participle they select indicates that they form a verb cluster at the second pole.

(5.15) a. Harry beseft wat er is$_1$ gebeurd$_2$ ...
Harry realises what there is happened ...
'Harry just realised what has happened ...' [LASSY, WR-P-E-I-0000004258.p.7.s.137]

b. ...toen ze voor de eerste keer rechtstreeks verkozen$_2$ werden$_1$.
...when they for the first time directly elected were
'...when they were directly elected for the first time' [CGN, fvf601267__55]

Worden 'be' always selects a past participle, but *hebben* and *zijn* also combine with an infinitive if that infinitive selects another verb, i.e. the IPP verbs.

(5.16) a. ...omdat ie het fotorolletje van een ondergeschikte heeft$_1$ laten$_2$
...because he the roll of film of a minor has let
verdwijnen$_3$.
disappear
'...because he has made the film of a minor disappear.' [CGN, fni007446__195]

b. en ik meen dat Ivo er daar toen is$_1$ gaan$_2$ halen$_3$.
and I mean that Ivo there there then is go get
'and I think that Ivo then went there to get some.' [CGN, fva400092__203]

Hebben 'have' does not appear as an IPP verb itself. *Zijn* 'be' does, but in that case it always appears in the form of *wezen*, which has an aspectual meaning, cf. (5.19a). The treebank results show that the verbs of the perfect and the passive are the most frequently occurring clustering verbs, as they account for 49.71% of the constructions.

The set of **modal verbs** accounts for 33.37% of the constructions in Table 5.7. It comprises the core modals *kunnen* 'can', *moeten* 'must', *mogen* 'may', and *zullen* 'shall', and the modals *hoeven* 'need', and *dienen* 'ought to'. The core modals all select a bare infinitive as their complement. Some examples are given in (5.17).

(5.17) a. Het lijkt erop dat de stad stilaan in de rook zal$_1$ verstikken$_2$.
it seems there-on that the city gradually in the smoke will suffocate
'It seems that the city will gradually suffocate in the smoke.' [LASSY, WR-P-P-B-0000000001.p.13.s.2]

b. Groot-Brittannië had het Suezkanaal kunnen$_1$ afsluiten$_2$, maar was bang
Great Britain had the Suez Canal can close but was afraid
voor een oorlog.
for a war
'Great Britain could have closed the Suez Canal, but was afraid of a war.' [LASSY, WR-P-E-I-0000051928.p.1.s.156]

c. Eind september werd bekend gemaakt dat hij 'het hart van Antwerpen'
end september was known made that he the heart of Antwerp
mag$_1$ ontwerpen$_2$ ...
may design ...
'At the end of September it was announced that he will design the 'heart of Antwerp' ...' [LASSY, dpc-cam-001283-nl-sen.p.11.s.5]

The verbs *hoeven* 'need' and *dienen* 'ought' typically select a *te*-infinitive, see section 5.3. In the case of *hoeven* the *te* can be dropped (5.18a). The corpus data contain an instance of *dienen* in combination with a past participle (5.18b), but in this case the entire *te*-infinitive (*te worden* 'to be') is dropped. This is indicated by the passive translation of the construction.

(5.18) a. nee ze heeft er niet veel moeite voor hoeven$_1$ doen$_2$.
no she has there not much effort for need do
'no she did not need to put a lot of effort in that.' [CGN, fna000446__412]

b. ... De verruiming van de vaargeul ..., waarvan de uitvoering dient$_1$
... the enlargement of the channel ..., of which the execution ought
gestart$_2$ vóór het einde van 2007 ...
started before the end of 2007 ...
'... the enlargement of the channel..., of which the execution should be started before the end of 2007 ...' [LASSY, WR-P-E-J-0000000018.p.8.s.1]

A third group of subject-oriented clustering verbs are the **aspectual verbs**, accounting for 8.81% of the constructions. They typically select an infinitive, but a subset (i.e. *komen* 'come', *staan* 'stand', *zitten* 'sit', and *liggen* 'lie') can also select a past participle. Some examples are given in (5.19).

(5.19) a. we waren even wat wezen$_1$ drinken$_2$ in de Oude Haven.
we were just something be drink in the Old Harbour
'we just went for a drink in the Old Harbour.' [CGN, fnc008063__58]

b. ja en je gaat ook liggen$_1$ wachten$_2$.
yeah and you go also lie wait
'yes and you will also be waiting.' [CGN, fna000938__136]

c. want ja je wou er vandaag toch over komen$_1$ praten$_2$?
because yeah you wanted there today actually about come talk
'because you wanted to come and talk about it today, right?' [CGN, fvf600243__110]

d. ... als Zijne Koninklijke Hoogheid statig The Mall op komt$_1$ gereden$_2$.
... if His Royal Highness solemnly The Mall on comes driven
'... if His Royal Highness solemnly drives up The Mall.' [LASSY, dpc-ind-001651-nl-sen.p.6.s.2]

Completing the list of clustering subject-oriented verbs are the **subject control verbs** *willen* 'want', *leren* 'learn', *durven* 'dare', and *proberen* 'try':

(5.20) a. Wie niet horen$_2$ wilde$_1$, moest voelen.
who not hear wanted must feel
'He who does not want to hear it, has to feel it. (proverb)' [LASSY, WR-P-P-I-0000000249.p.4.s.6]

b. maar ik wou echt wel piano leren$_1$ spelen$_2$.
but I wanted really rather piano learn play
'but I really wanted to play the piano.' [CGN, fva400073__35]

c. maar 'k heb vanmorgen nog proberen$_1$ bellen$_2$ naar Lutje en daar nam
but I have this morning still try call to Lutje and there picked
niemand op dus ...
no one up so ...
'but I have tried to call Lutje this morning and no one picked up the phone so...' [CGN, fvc701065__48]

The set of subject-control verb account for 4.46% of the constructions in Table 5.7, i.e. the smallest category of the subject-oriented verbs.

As will be shown in section 5.3, most of the aspectual verbs, as well as *leren*, *durven*, and *proberen* can also select a *te*-infinitive. The treebank results in Table 5.7 are constructions where the infinitival marker *te* is omitted, which is an indication of their status as a clustering verb (cf. section 1.8).

Except for the perfect and passive auxiliaries, the subject-oriented verbs in Table 5.7 obligatorily or optionally appear as IPP in the perfect tense. The relation between the IPP effect and cluster formation will be discussed in more detail in sections 5.3 and 5.4.

Object-oriented clustering verbs

The set of object-oriented verbs listed in Table 5.7 are all object raisers. Similarly to the subject-oriented verbs, the majority invariably selects a bare infinitive as its verbal complement. The results indicate that in contrast to the subject-oriented verbs, the object-oriented verbs occur less frequently as a selector in verb clusters.

The object-oriented verbs that occur most frequently in the treebank data as VC-selectors are the **causative verbs** *laten* 'make' and *doen* 'do'. They account for 2.64% of the constructions in Table 5.7.

(5.21) a. kan 'k misschien straks nog wel laten$_1$ zien$_2$.
can I maybe later still rather let see
'I can maybe show that later.' [CGN, fna000372__256]

b. het zijn deze bedrijven die ook de boerenstiel ten onder doen$_1$ gaan$_2$.
it are these companies that also the farmer's trade to under do go
'It is these companies that make the farmer's trade go down.' [CGN, fvg600014__97]

A second set of clustering object-oriented verbs are the **perception verbs** *zien* 'see', *horen* 'hear', and *voelen* 'feel'. They only account for 0.56% of the constructions in Table 5.7. Those verbs canonically select a bare infinitive, but there are a few cases in which *zien* selects a participle. Similar to the combination of *dienen* with a participle in (5.18b), those are passive constructions. If one adds the infinitive *worden* 'be', the meaning of the sentence does not change.

(5.22) a. Ik heb mensen al horen$_1$ beweren$_2$ dat ze het niet erg vinden om
I have people already hear claim that they it not bad find to
belastingen te betalen ...
taxes to pay ...
I have heard people claim that they do not mind paying taxes ... [LASSY, dpc-ind-001634-nl-sen.p.10.s.5]

b. Het maakt bijvoorbeeld verschil of iemand zijn aanvallen voelt$_1$
It makes for example difference whether someone his attacks feels
aankomen$_2$ en ook hoe die aanvallen verlopen.
come and also how those attacks evolve.
For example it makes a difference whether someone feels the attacks coming and also how those attacks evolve. [LASSY, WR-P-P-C-0000000055.txt-55]

c. De vacaturesite Monsterboard.nl heeft het aantal ICT-vacatures in drie
The job site Monsterboard.nl has the amount ICT vacancies in three
maanden tijd zien$_1$ verdubbelen$_2$.
months time see duplicate.
'The job site Monsterboard.nl has seen the amount of ICT vacancies duplicate in three months time.' [LASSY, WS-U-E-A-0000000233.p.14.s.5]

d. Pronk hoopt op een voortzetting van de paarse coalitie van PvdA, VVD en
Pronk hopes for a continuation of the purple coalition of PvdA, VVD and
D66, maar wil die het liefst aangevuld$_2$ zien$_1$ met
D66, but wants that the preferable.SUP complemented see with
GroenLinks.
GroenLinks.
'Pronk is hoping for a continuation of the purple coalition of PvdA, VVD and D66, but preferably wants it to be complemented by GroenLinks.' [LASSY, WR-P-P-H-0000000054.p.5.s.1]

A third set of object-oriented verbs contains the **benefactives** *helpen* 'help' and *leren* 'teach'. Accounting for a meagre 0.05% of the constructions in Table 5.7, their frequency is also very low compared to the other types of object-oriented verbs.

(5.23) a. ... een Stichting ..., die de democratie en de rechtstaat in
... a Foundation ..., which the democracy and the constitutional state in
Congo wil$_1$ helpen$_2$ bevorderen$_3$.
Congo wants help stimulate
'... a Foundation ..., which wants to stimulate the democracy and the constitutional state in Congo.' [LASSY, WR-P-P-I-0000000148.p.9.s.5]

b. 'k heb mijn kleine kinderen daar ook leren$_1$ zwemmen$_2$.
I have my little children there also teach swim.
'I have taught my young children to swim there as well.' [CGN, fva400659__44]

The causative, perception and benefactive verbs all obligatorily (causatives, perception verbs) or optionally (benefactives) appear as IPP in the perfect, cf. section 5.4. In addition, the class of object-oriented verbs contains three less typical clustering verbs that do not appear as IPP: *vinden* 'find, think', *hebben* 'have', and *krijgen* 'get'.

If they appear as a selector of a VC, *vinden* and object-oriented *hebben* always select a bare infinitive:

(5.24) a. Zo sprak ik een vrouw die het niet vond$_1$ kunnen$_2$ dat haar Xhosa
so spoke I a woman that it not found can that her Xhosa
buren vuurtjes maakten in de tuin.
neighbours fires.DIM made in the garden.
'So I spoke to a woman who could not stand her Xhosa neighbours making fires in the garden.' [LASSY, dpc-cam-001018-nl-sen.p.8.s.5]

b. ... want we hebben dus één plek waar we dus erg veel
... because we have thus one spot where we thus very much
nachtdieren bij elkaar hebben$_1$ zitten$_2$.
nocturnal animals with each other have sit
'... because we have one spot where a lot of nocturnal animals are put together.' [CGN, fnj007495__125]

c. omdat ik dan daarover wel wat in m'n kast steken$_2$ heb$_1$.
because I then there-about rather something in my cupboard put have
'because I have something about that lying in my cupboard' [CGN, fvb400117__63]

In contrast to the subject-oriented perfect auxiliary *hebben*, the use of *hebben* in (5.24b–c) is closer to the use of *hebben* as a main verb, i.e. in the sense of 'own, possess'. If one would drop the infinitival complement, the meaning of the sentence does not change and the sentence would remain well-formed.

Constructions like (5.24) can hardly be put in the perfect tense, and thus appear not as IPP in the treebanks. Their word order suggests that they are clustering verbs, however: *vinden* always selects its infinitival complement to the right in the treebank data; *hebben* shows some word order variation in the cluster, but it typically selects its complement to the right. In colloquial Dutch, *hebben* can be put in the perfect tense, but in that case it has the form of a past participle. No treebank instances were found, but (5.25) is an example from the Internet:

(5.25) ... al heb ik er ooit ook wel eens eentje gewoon los op
... although have I there once also rather sometime one.DIM just loose on
de bodem van mijn kast liggen gehad.
the bottom of my cupboard lie had
'although I once had one lying loose on the bottom of my cupboard.' [Google.be, 7-7-2015]

The final verb in the list of object-oriented verbs is *krijgen* 'get'. It often serves as a passive auxiliary verb in constructions with ditransitive verbs.[14] If it selects a passive participle, the word order possibilities in (5.26) suggest that it is a clustering verb:

[14]In constructions with the *krijgen* passive, the indirect object of the ditransitive verb acts as the

(5.26) a. Ze vinden het opmerkelijk ... dat de BBC de zwarte piet krijgt$_1$
they find it remarkable ... that the BBC the jack of spades gets
toegespeeld$_2$.
passed
'They find it remarkable ... that the BBC got blamed.' [LASSY, WS-U-E-A-0000000028.p.20.s.3]

b. ... terwijl het naburige Lierde de meer bescheiden titel van
... while the neighbouring Lierde the more modest title of
mattentaartengemeente toegewezen$_2$ kreeg$_1$.
mattentaarten municipality assigned got
'... while the neighbouring town Lierde received the more modest title of 'mattentaarten municipality'.' [LASSY, WR-P-E-I-0000039589.p.11.s.1]

In contrast to the other clustering verbs, however, *krijgen* canonically selects its complement to the left. This is observed by Haeseryn et al. (1997) and confirmed by the treebank data: Out of the 38 constructions with *krijgen*, it selects its complement to the right in only 10 cases. Moreover, if constructions with *krijgen* appear in the perfect tense, it always appears as a past participle and not as an IPP verb. In those constructions, the participle always appears in front of its selector, as in (5.27).[15] Those characteristics are typical for predicative selecting verbs, as motivated in Van Eynde (2015), and in the examples with *raken* 'get, become' in section 1.10.

(5.27) ... maar als je kijkt naar het lijstje wat we onder andere vanmiddag
... but if you look at the list.DIM what we amongst others this afternoon
bij brief van de van de van de van de VNG overhandigd$_3$ hebben$_1$ gekregen$_2$
with letter of the of the of the of the VNG over-handed have got
dan ...
then ...
'but if you look at the list that we have received with the letter of the VNG then...'
[CGN, fng000175__99]

The examples in (5.26) and (5.27) show that it is hard to classify *krijgen* as either a clustering verb or a predicate selector. The word order variation observed in (5.26) as well as its occurrence as an auxiliary in the *krijgen* passive suggests that it should be included in the list of clustering verbs.

---

subject of the passive construction,e.g. *de BBC* 'the BBC' in (5.26a), whereas in regular passives, the direct object acts as the subject of the construction, e.g. *zij* 'they' in (5.15b). For an HPSG analysis of the *krijgen* passive, see Kordoni & van Noord (2009).

[15]This example is not part of the set of constructions listed in Table 5.7.

Non-finite predicative complements

In 63 constructions a past participle is selected as a predicative complement, rather than as a verbal complement. In contrast to clustering verbs, which typically select their complement to the right, predicate selectors invariably select their complement to the left. Some examples are given in (5.28).

(5.28) a. uhm ook omdat ik mij daartoe een beetje verplicht voel naar de leerlingen
uh also because I me there-to a little obliged feel to the pupils
toe ...
to ...
'uh also because I feel somewhat obliged towards the pupils about that...'
[CGN, fvb400155__421]

b. de aanvoer van mest moet immers verzekerd blijven.
the supply of manure must after all guaranteed remain
'after all the manure supply must be guaranteed.' [CGN, fvg600014__57]

The examples in (5.28a–b) show that there are a number of verbs (a.o. *voelen* 'feel' and *blijven* 'remain') that can appear as predicate selectors (if they select a participle) and as VC-selectors (if they select a bare infinitive).

5.2.4 Conclusion

This section has presented an overview of the bare infinitival and participial verb clusters found in the treebanks. The general observation with regard to the word order variation is that the position of the past participle in the cluster is free, while the other elements in the cluster tend to appear in the ascending word order, confirming the generalizations of Broekhuis & Corver (2015).

Besides investigating the cluster types and the word order variation in the cluster, the list of clustering verbs selecting an infinitival and/or a participial VC was extracted from the data. This resulted in a first set of 32 different clustering verbs. The verbs listed in Table 5.7 are also mentioned in Haeseryn et al. (1997: 1077–1081). What that list does not contain, however, is the division into larger verb categories that can appear as a clustering verb. Moreover, the frequency results obtained from the treebanks give an indication which verbs are typical clustering verbs, and which verbs hardly occur in verb clusters.

The typology in Figure 5.2 provides an overview of the different verb categories that select a bare infinitive and/or a past participle in the cluster, as well as their relative frequency in the treebanks.

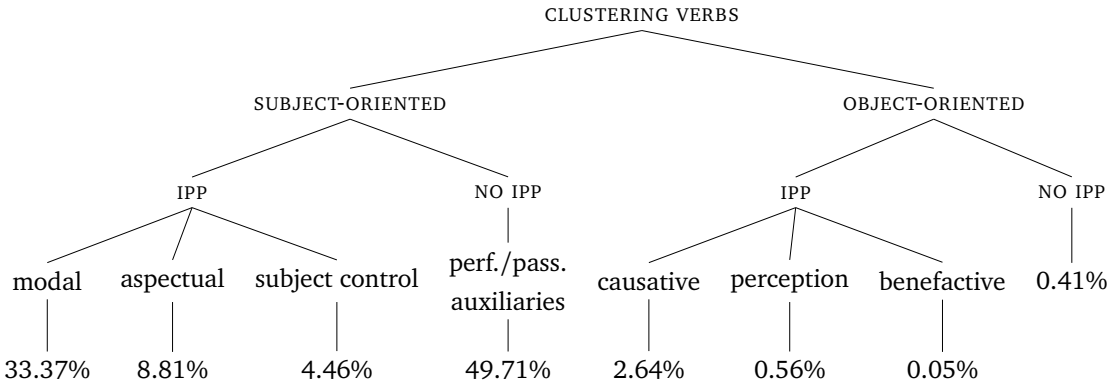


Figure 5.2: Treebank-based typology of clustering verbs selecting an infinitival and/or a participial complement

Except for the auxiliaries of the perfect and the passive, *krijgen*, *vinden*, and object-oriented *hebben*, the verbs listed in Table 5.7 are all IPP verbs (if they appear in the perfect tense), indicating that the ability to occur as IPP is an important diagnostic to identify the Dutch clustering verbs.[16] Verbs that do not show the IPP effect, can be identified as clustering verbs by means of the word order variation in the cluster.

5.3 Verb clusters with a *te*-infinitive

In chapter 1 it was argued that constructions with *te*-infinitives can be split up into constructions where the *te*-infinitive is part of the cluster, as in (5.29a), and constructions in which it is not, as in (5.29b–c).

(5.29) a. 'k ben blij dat ik zo veel belangstelling *heb weten te wekken*.
I am happy that I so much interest have known to raise
'I am glad that I have been able to raise so much interest.' [CGN, fnf007126__142]

[16] See also the descriptive study in section 1.4 and the treebank-based investigation of IPP verbs in section 5.4.

b. en ik denk dat men daarin *moet trachten* het juiste evenwicht te zoeken.
and I think that one there-in has try the right balance to search
'and I think that one has to try to find the right balance in that.' [CGN, fvg600012__38]

c. ... ik denk dat het voor een stuk moeilijk te vermijden *is* ...
... I think that it for a piece difficult to avoid is
'... partly I think that it is hard to avoid' [CGN, fvj601108__6]

In constructions with IPP, such as (5.29a), the *te*-infinitive is canonically considered to be part of the verb cluster. In (5.29b) the cluster consists of a finite verb and a bare infinitive (*moet trachten* 'have try'), whereas the *te*-infinitive is in the Nachfeld. In (5.29c) the *te*-infinitive is a predicative infinitive which is in the Mittelfeld, rather than in the second pole.

While the position of the *te*-infinitive is relatively straightforward in the examples in (5.29), it is more difficult classify the constructions in (5.30), i.e. cases of optional IPP verbs that do not occur as IPP.

(5.30) a. ja dan uh als je de grote lijnen 'ns probeert te schetsen...
yeah then uh if you the big lines once try.FIN to sketch.INF
'yeah and then if you try to sketch the broad outlines ...' [CGN, fnf007116__4]

b. oh ze zal proberen wakker te worden zegt ze.
oh she will.FIN try.INF awake to become.INF says she
'oh she says she will try to be awake.' [CGN, fnc008001__218]

Based on the verb form and the position of the arguments it cannot be decided whether (5.30a) is an instance of the third construction or of a verb cluster, and whether *wakker* 'awake' in (5.30b) is a cluster creeper, or in the Nachfeld with the *te*-infinitive, cf. the discussion in section 1.10.

Since constructions with a *te*-infinitive in the Nachfeld or in the Mittelfeld are tagged similarly to clustering constructions (i.e. both constructions receive a VC tag in the treebanks), it is more complex to detect the clusters with *te*-infinitives in the treebanks compared to the extraction of clusters with bare infinitives and/or a past participle. Therefore, all constructions in which a *te*-infinitive is selected by a finite verb or an infinitive are extracted from the treebanks (section 5.3.1). Next, the constructions are classified according to the word order of the verbs and the type of verb forms occurring in the constructions (section 5.3.2). In section 5.3.3 the set of clustering verbs is identified by considering the IPP status of the selecting verbs and their

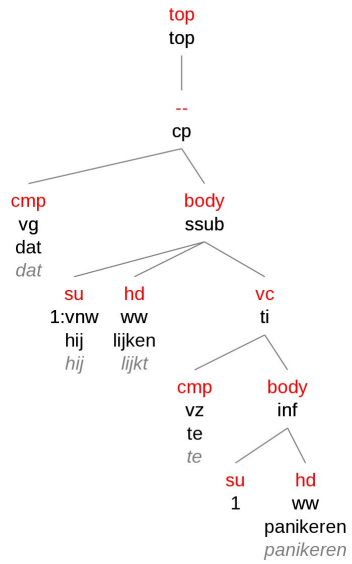
adjacency to the verbal complement. The conclusions of this section are summarized in section 5.3.4.

5.3.1 Extracting the constructions

The constructions with a *te*-infinitive can be extracted in a similar way to the participial and bare infinitival clusters. For instance, verb-final constructions in which the finite verb selects a *te*-infinitive, as in (5.31), (e.g. *dat hij lijkt te panikeren* 'that he seems to be panicking'), can be retrieved by the query in (5.32).

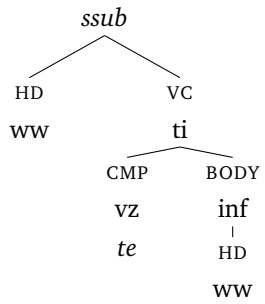
(5.31) a. ... dat hij lijkt te panikeren.
... that he seems to panic
'... that he seems to be panicking.'

b.



(5.32) a.
```
//node[@cat="ssub" and
        node[@rel="hd" and @pt="ww" and @wvorm="pv"] and
        node[@rel="vc" and @cat="ti" and
            node[@rel="cmp" and @pt="vz" and @lemma="te"] and
            node[@rel="body" and @cat="inf" and
                node[@rel="hd" and @pt="ww"] ] ] ]
```

b.



Due to the attachment of the infinitival marker *te*, *te*-infinitives have a more complex structure in the treebanks compared to bare infinitives and participles, as illustrated in the tree representation (5.32b). As a consequence, the extra level of embedding results in longer and more complex search instructions.

Similar to the queries in section 5.2, one can control the number of verbs embedded under the *te*-infinitive by means of the `not()`-operator. For instance, the query in (5.33) returns verb-initial constructions in which there is one (and at most one) bare infinitive embedded under the *te*-infinitive node, returning constructions of the type *Hij lijkt morgen te kunnen komen* 'He seems to be able to come tomorrow'.

(5.33)
```
//node[(@cat="smain" or @cat="sv1") and
      node[@rel="hd" and @pt="ww" and @wvorm="pv"] and
      node[@rel="vc" and @cat="ti" and
          node[@rel="cmp" and @pt="vz"] and
          node[@rel="body" and @cat="inf" and
              node[@rel="hd" and @pt="ww"] and
              node[@rel="vc" and @cat="inf" and
                  node[@rel="hd" and @pt="ww"] and
                  not(node[@rel="vc" and (@cat="ti" or @cat="inf"
                  or @cat="ppart" or @pt="ww")]) ] ] ] ]
```

Other constructions with *te*-infinitives can be found by variations of the query in (5.33).

5.3.2 Cluster types and word order variation

The treebanks only contain 1,387 constructions with a *te*-infinitive which potentially contain a verb cluster. They are another subset (1.78%) of the 78,110 constructions in which a verb selects a VC (see Table 5.1 in section 5.1). Compared to the ca.

21,000 infinitival and participial clusters discussed in section 5.2, it is a rather small set. Table 5.8 gives an overview of the constructions according to their length. Also in this case, no constructions with more than four verbs were found. Similar to the results in section 5.2.2, the figures are the sum of the queries for non-terminal and terminal VCs. The distribution of the constructions with respect to the amount of verbs is very similar to the figures in Table 5.3, i.e. around 90% for the two-verb constructions, around 9% for the three-verb constructions, and less than 1% for the four-verb constructions.

	CGN		LASSY		**Sum**	
	CONSTRUCTIONS	%	CONSTRUCTIONS	%	CONSTRUCTIONS	%
2 VERBS	489	92.6	765	89.1	1,254	90.4
3 VERBS	35	6.6	92	10.7	127	9.2
4 VERBS	4	0.8	2	0.2	6	0.4
Sum	**528**	**100**	**859**	**100**	**1,387**	**100**

Table 5.8: Length of the constructions with a *te*-infinitive

It will be shown in this and the following section that only a subset of the results presented in Table 5.8 are actually verb clusters with a *te*-infinitive. In order to identify those constructions, the remainder of this section will zoom in on the constructions in a similar fashion as was done in section 5.2.2. The constructions in Table 5.8 will be discussed according to their word order variation and the word form of the verbs.

Tables 5.9 to 5.11 present the results per cluster length. The order of selection and the word form of the verbs is indicated in the rows, while the linear order is given in the columns.

The majority of the two-verb constructions appear in the canonical 1-2 order, but the results in Table 5.9 show that both treebanks also contain a considerable amount of 2-1 constructions. While this was expected for the constructions with a past participle, this is somewhat surprising for the infinitival complements, given the generalizations from Broekhuis & Corver (2015).

The constructions in which a *te*-infinitive selects a past participle or a bare infinitive are all clustering (i.e. the constructions at the top of Table 5.9). Some examples are given in (5.34). In those constructions, the verb that selects the *te*-infinitive occupies the first pole.

	CGN		LASSY		**Sum**
	1-2	2-1	1-2	2-1	
TE INF_1 PSP_2	14	16	81	46	157
TE INF_1 INF_2	27	0	69	0	96
$FINITE_1$ TE INF_2	250	93	343	127	813
INF_1 TE INF_2	77	12	80	13	182
TE INF_1 TE INF_2	0	0	5	1	6
Sum	**368**	**121**	**578**	**187**	**1,254**

Table 5.9: Two-verb constructions with a *te*-infinitive in CGN and LASSY Small

(5.34) a. de Amerikaanse zangeres lijkt te veel hooi op haar vork te hebben$_1$
the American singer seems too much hay on her fork to have
genomen$_2$.
taken
The American singer seems to have taken on too much.' [CGN, fnk005898__2]

b. De bladen zeggen niet op het aanbod ingegaan$_2$ te zijn$_1$.
the magazines say not on the offer in-gone to be
'The magazines say they have not taken up the offer.' [LASSY, WS-U-E-A-0000000223.p.51.s.2]

c. het bleek ne man te moeten$_1$ zijn$_2$.
it turned out a man to must be
It turned out that it had to be a man.' [CGN, fvc901001__7]

The selecting verbs are all obligatory clustering verbs that were already listed in Table 5.7. The word order in those constructions is also similar to the constructions discussed in section 5.2: The participle can occur before or after its selector (5.34a–b), whereas a bare infinitive invariably follows its selector (5.34c).

The constructions in which the selector of the *te*-infinitive is in verb-final position show both the 1-2 and the 2-1 order. Some examples are given in (5.35).

(5.35) a. Ik was erg tevreden dat er in Leuven plaats voor bleek$_1$ te zijn$_2$.
I was very pleased that there in Leuven place for turned out to be
'I was very pleased that there turned out to be a place for it in Leuven.' [dpc-cam-001020-nl-sen.p.12.s.6]

b. ... ik heb geen sleutel onder de mat hoeven$_1$ te stoppen$_2$ ...
... I have no key under the mat need to stop ...
'... I did not need to put a key under the mat...' [LASSY, dpc-rou-000479-nl-sen.p.10.s.14]

c. 'k denk dat er veel aan te doen$_2$ is$_1$ hè?
I think that there much on to do is right
'I think there is a lot to be done about it, right?' [CGN, fva400187__131]

d. Mojo Concerts zegt niets te verbergen$_2$ te hebben$_1$.
Mojo Concerts says nothing to hide to have
'Mojo Concerts says that they have nothing to hide.' [WS-U-E-A-0000000048.p.25.s.5]

e. ...je kan wel te weten$_2$ komen$_1$ welke kleuren welke uh hoogte voorstellen ...
...you can rather to know come which colours which uh height represent ...
'...you can find out which colours represent which height ...' [CGN, fvh400057__8]

Blijken 'turn out' and *hoeven* 'need' select a verbal complement in (5.35a–b), but this is not the case for *zijn* 'be' and *hebben* 'have' in (5.35c–d). As motivated in section 1.10, predicative infinitives such as *te doen* 'to do' and *te verbergen* 'to see' can be differentiated from verbal complements by the fact that they canonically appear in front of their selector and by their passive modal meaning. They obtain a position in the Mittelfeld, rather than in the second pole. Similarly, *te weten* 'to know' in (5.35e) is selected in the Mittelfeld as part of the fixed expression *te weten komen* 'find out' rather than as a verbal complement of *komen* 'come'. In contrast to the ascending constructions, none of the 2-1 constructions in which a *te*-infinitive appears to the left of its selector are instances of verb clusters.

With respect to the 1-2 constructions in which a finite verb, a bare infinitive or a *te*-infinitive select a *te*-infinitive, the selected *te*-infinitive can be in the cluster, as in (5.35a–b) and (5.36a), but it may also be in the Nachfeld, as in (5.36b), or the position of the *te*-infinitive might be ambiguous, as in (5.30a), repeated as (5.36c).

(5.36) a. Ik heb mijn agenda niet hoeven$_1$ om te gooien$_2$ ...
I have my agenda not need over to throw ...
'I did not have to completely change my schedule ...' [LASSY, dpc-rou-000479-nl-sen.p.10.s.14]

b. ...hoe ze zelf denkt$_1$ de macht weer naar zich toe te trekken$_2$ ...
...how she herself thinks the power again to herself to to draw ...
'...how she thinks to gain power for herself again...' [CGN, fnf007107__25]

c. ja dan uh als je de grote lijnen 'ns probeert$_1$ te schetsen$_2$ ...
yeah then uh if you the big lines once tries to sketch
'yeah and then if you try to sketch the broad outlines ...' [CGN, fnf007116__4]

The separable verb particle *om* in (5.36a) can be considered a cluster creeper (cf. section 5.5), while this is not the case for the multiple non-verbal elements occurring between *denkt* 'thinks' and *te trekken* 'to pull' in (5.36b). As discussed in section 1.10, optional IPP verbs like *probeert* 'tries' in (5.36c) are often ambiguous between clustering and non-clustering if they appear as a finite verb. The distinction between verbs selecting a *te*-infinitive in the second pole or in the Nachfeld will be discussed in further detail in section 5.3.3.

The observations for the three- and four-verb constructions are very similar to the observations for the two-verb constructions.

	CGN			LASSY			**Sum**
	1-2-3	1-3-2	3-1-2	1-2-3	1-3-2	3-1-2	
FINITE$_1$ TE INF$_2$ PSP$_3$	2	1	3	18	5	11	40
FINITE$_1$ TE INF$_2$ INF$_3$	7	0	0	10	0	0	17
INF$_1$ INF$_2$ TE INF$_3$	3	0	0	9	0	0	12
INF$_1$ TE INF$_2$ INF$_3$	1	0	0	2	0	0	3
TE INF$_1$ INF$_2$ INF$_3$	1	0	0	1	0	0	2
TE INF$_1$ INF$_2$ PSP$_3$	0	0	0	1	0	1	2
INF$_1$ TE INF$_2$ PSP$_3$	0	0	0	1	0	0	1
FINITE$_1$ INF$_2$ TE INF$_3$	13	0	3	30	0	2	48
FINITE$_1$ TE INF$_2$ TE INF$_3$	0	1	0	1	0	0	2
Sum	**27**	**2**	**6**	**73**	**5**	**14**	**127**

Table 5.10: Three-verb constructions with a *te*-infinitive in CGN and LASSY Small

Table 5.10 indicates that the position of past participles is variable, as exemplified in (5.37a–c), while bare infinitives invariably follow their selector (5.37d). Also in the constructions with three verbs, verbs selecting a bare infinitive or a past participle occur in the list of clustering verbs in Table 5.7.

(5.37) a. ...en lijken onder de Hunnen te moeten$_1$ worden$_2$ gerekend$_3$.
...and seem under the Huns to must be counted
'...and seem to be considered as Huns.'[LASSY, WR-P-E-I-0000027216.p.1.s.32]

b. . . . dat de bedragen hierboven met een factor 5,5 dienen$_1$ vermenigvuldigd$_3$
. . . that the amounts above with a factor 5.5 need multiplied
te worden$_2$.
to be
'. . . that the amounts above need to be multiplied by a factor of 5.5.' [LASSY, dpc-bmm-001089-nl-sen.p.14.s.3]

c. . . . maar bleek deze voorkomen$_3$ te kunnen$_1$ worden$_2$ . . .
. . . but turned out this avoided to can be . . .
'. . . but it turned out to be possible to avoid it . . .' [LASSY, WR-P-E-I-00000368 34.p.1.s.122]

d. ik meen ook te hebben$_1$ mogen$_2$ begrijpen$_3$ dat . . .
I mean also to have may understand that . . .
'I also think to have understood that . . .' [CGN, fvg600012__10]

The *te*-infinitives canonically follow their selector, but as in the 2-1 constructions, there are a few instances in which the *te*-infinitive precedes its selector if it is predicative or part of a fixed expression, i.e. the constructions at the bottom of Table 5.10. Some examples are given in (5.38).

(5.38) a. . . . dat de verschillen te wijten$_3$ kunnen$_1$ zijn$_2$ aan het eiwit haptoglobine.
. . . that the differences to attribute can be on the protein haptoglobin
'. . . that the differences can be attributed to the protein haptoglobin.' [LASSY, WR-P-P-I-0000000100.p.4.s.2]

b. . . . omdat ze dacht$_1$ met een everzwijn te maken$_3$ te hebben$_2$.
. . . because she thought with a wild boar to make to have
'. . . because she thought to be facing a wild boar.' [CGN, fvf600049__5]

In the constructions with the ascending 1-2-3 order, the *te*-infinitive might be selected in the second pole, in the Nachfeld, or the construction might be ambiguous with respect to the position of the *te*-infinitive, cf. the examples in (5.29a–b) and (5.30).

The treebanks only contain six four-verb constructions with a *te*-infinitive, cf. Table 5.11.

	CGN		LASSY	**Sum**
	1-2-3-4	1-2-4-3	1-2-3-4	
INF_1 INF_2 TE INF_3 PSP_4	0	1	1	2
FINITE_1 TE INF_2 INF_3 PSP_4	0	0	1	1
FINITE_1 INF_2 INF_3 TE INF_4	1	0	0	1
INF_1 INF_2 TE INF_3 INF_4	1	0	0	1
FINITE_1 TE INF_2 TE INF_3 INF_4	1	0	0	1
Sum	**3**	**1**	**2**	**6**

Table 5.11: Four-verb constructions with a *te*-infinitive in CGN and LASSY Small

Five out of the six examples occur in the ascending order, as in (5.39a). The example with the 1-2-4-3 order is presented in (5.39b).

(5.39) a. Ricardo Oliveira, die als pasmunt leek_1 te zullen_2 worden_3 gebruikt_4 in de
Ricardo Oliveira who as change seemed to will be used in the
transfer van Ronaldo ...
transfer of Ronaldo ...
'Ricardo Oliveira, who seemed to be used as bargaining chip in the transfer of Ronaldo' [LASSY, dpc-rou-000349-nl-sen.p.11.s.2]

b. het CDA had eerder al laten_1 weten_2 onthutst_4 te zijn_3.
the CDA had earlier already let know upset to be
'CDA had already announced earlier to be upset.' [CGN, fnk001946__5]

(5.39a) is an example of a verb cluster with four verbs. In (5.39b), *laten* 'let' selects *weten* 'know', which is a main verb in this construction. The past participle and the *te*-infinitive occur in the Nachfeld.

Summing up, the treebank data discussed in this section show that the set of constructions containing a *te*-infinitive is very diverse. It turns out that verbs selecting a bare infinitive or a past participle were already identified as obligatory clustering verbs in section 5.2 (cf. Table 5.7).

While past participles can occur before or after their selecting verb, bare infinitives canonically follow their selector. With respect to the *te*-infinitives, a distinction was made between *te*-infinitives that are selected as a verbal complement versus predicative *te*-infinitives and *te*-infinitives that are a part of an idiomatic expression. The former appear to the right of the selecting verb, whereas the latter canonically appear to the left of their selector. The verbs that select a *te*-infinitive as a verbal complement will be discussed in further detail in section 5.3.3.

5.3.3 Clustering verbs

This section investigates which verbs select a *te*-infinitive in the second pole, i.e. as a part of the verb cluster, and which verbs select a *te*-infinitive in the Nachfeld. Table 5.12 lists the constructions in which the selector of the *te*-infinitive is verb-final.[17] For each construction, it is indicated in how many cases the selector is adjacent to the *te*-infinitive (ADJ), and how often it is separated from the *te*-infinitive by one or more other elements (NOT ADJ). The constructions in which the *te*-infinitive appears in front of its selector are left out, as it was shown in section 5.3.2 that the *te*-infinitive is not a verbal complement in those constructions .

	CGN		LASSY		**Sum**
	ADJ	NOT ADJ	ADJ	NOT ADJ	
FINITE_1 TE INF_2 PSP_3	6	0	31	3	40
INF_1 INF_2 TE INF_3 PSP_4	0	1	0	1	2
INF_1 TE INF_2 PSP_3	0	0	0	1	1
FINITE_1 TE INF_2 INF_3 PSP_4	0	0	1	0	1
FINITE_1 TE INF_2	190	60	201	142	593
INF_1 TE INF_2	41	36	33	47	157
FINITE_1 INF_2 TE INF_3	7	6	13	17	43
FINITE_1 TE INF_2 INF_3	4	3	2	8	17
INF_1 INF_2 TE INF_3	0	3	2	7	12
TE INF_1 TE INF_2	0	0	0	5	5
INF_1 TE INF_2 INF_3	0	1	0	2	3
FINITE_1 INF_2 INF_3 TE INF_4	0	1	0	0	1
INF_1 INF_2 TE INF_3 INF_4	0	1	0	0	1
FINITE_1 TE INF_2 TE INF_3	0	0	0	1	1
FINITE_1 TE INF_2 TE INF_3 INF_4	0	1	0	0	1
Sum	**248**	**113**	**283**	**234**	**878**

Table 5.12: Constructions in which a *te*-infinitive is selected in the second pole or in the Nachfeld

[17] As it cannot be determined by the position nor the verb form of the verb in the first pole whether it is clustering or not, the constructions in which the selector of the *te*-infinitive is verb-initial are left out.

As motivated in chapter 1, the ability of the selecting verb to appear in IPP constructions is an important diagnostic for clustering. The treebank data confirm this observation. If we compare the constructions in Table 5.12 with respect to the ability of the selecting verb to appear as IPP,[18] we get the figures in Table 5.13.[19]

Selector	**Adjacent**	**Not adjacent**	**Sum**
OBLIGATORY IPP	331	43	374
OPTIONAL IPP	164	147	311
NO IPP	37	158	195
Sum	**532**	**348**	**880**

Table 5.13: IPP versus adjacency for verbs selecting a *te*-infinitive

The **obligatory IPP verbs** appear adjacent to their verbal complement in 88.5% of the cases (i.e. in 331 constructions). If we take a closer look at the 43 non-adjacent constructions, the majority (30 constructions) turn out to be instances in which the separable verb particle occurs between the verbs, as in (5.40a). One instance with a direct object was found (5.40b), and the remaining 12 cases are instances where the verbs are separated by interruptions (e.g. by *uh* in the spoken data).

(5.40) a. ...winst die overigens bij voorbaat leek$_1$ vast te staan$_2$.
...profit which anyway by advance seemed fixed to stand
'...profit that seemed to be set in advance.' [WS-U-E-A-0000000051.p.25.s.2]

b. ...en waar dat wij stonden$_1$ handjes te schudden$_2$ hè.
...and where that we stood hands.DIM to shake right
'...and where we were shaking hands, right.' [CGN, fva400282__86]

In contrast to the obligatory IPP verbs, the set of **optional IPP verbs** shows a more equal distribution between adjacent and non-adjacent constructions. If the selecting verb occurs as IPP (i.e. in 16 out of the 311 hits), the verbs are typically adjacent (12 hits), as in (5.41a), but instances with a cluster creeper (4 hits) occur as well (5.41b).

[18] The IPP status of the selecting verbs was verified using examples from the treebanks and the Web, cf. section 5.4.

[19] The FINITE$_1$ TE INF$_2$ TE INF$_3$ and the FINITE$_1$ TE INF$_2$ TE INF$_3$ INF$_4$ constructions are counted twice, as those constructions contain two verbs selecting a *te*-infinitive. Hence, the constructions in Table 5.13 add up to 880.

(5.41) a. De Israëlische linkerzijde ... was weer een vredeskamp beginnen$_1$ te vormen$_2$.
the Israeli left side ... was again a peace camp start to form
'The Israeli left wing ... had again started to set up a peace camp.' [LASSY, WR-P-P-I-0000000204.p.9.s.4]

b. buitenlanders hebben dit ook proberen$_1$ uit te leggen$_2$ ...
foreigners have this also try out to explain
'Foreigners also have tried to explain this ...' [CGN, fnj007353__60]

No instances were found in which full phrases or multiple complements occur between the IPP and the *te*-infinitive, which indicates that optional IPP verbs are always clustering if they appear as IPP. This is not the case if they do not appear as IPP (i.e. in 295 out of the 311 hits).

There are 152 out of the 164 adjacent constructions with an optional IPP verb in which the selecting verb does not appear in the IPP form. In 131 cases all arguments of the *te*-infinitive are in the Mittelfeld. The construction with *proberen* 'try' in (5.30a) is an example of this set. In 21 examples the *te*-infinitive does not select any non-subject dependants, which causes it to be adjacent to its selector (5.42).

(5.42) en als jouw geweten begint$_1$ te spreken$_2$ hè ...
and if your conscience begins to speak right ...
'and if your conscience starts to speak, right' [CGN, fnh009109__147]

As the optional IPP verbs can appear in IPP constructions and in the third construction, it is not possible to classify constructions like (5.42) as either clustering or non-clustering.

There are 143 constructions with an optional IPP verb in which the selecting verb does not appear in the IPP form and in which the selector is not adjacent to the *te*-infinitive. 83 constructions are clearly non-clustering, as all non-subject arguments belonging to the *te*-infinitive are in the Nachfeld. An example is given in (5.43).

(5.43) ... omdat de provincie weigert$_1$ extra geld aan de gemeente te betalen$_2$.
... because the province refuses extra money to the municipality to pay
'... because the province refuses to pay extra money to the municipality' [LASSY, WS-U-E-A-0000000015.p.32.s.11]

The remaining examples are also difficult to classify as either clustering or non-clustering. In 44 cases a separable particle occurs between the verbs, while the other

arguments (if any) obtain a position in the Mittelfeld. (5.30b) is an example of this set. Another example is given in (5.44).

(5.44) Toen de Noormannen ... begonnen$_1$ binnen te vallen$_2$ ...
when the Norsemen ... began inside to fall ...
'When the Norsemen ... started to invade' [LASSY, WR-P-E-I-0000044854.p.3.s.7]

While the examples with separable verb particles occurring between the verbs are again cases in which it is hard to decide whether we are dealing with clusters or non-clusters, there is a small set of examples (16 hits) indicating that optionally IPP verbs are not clustering if they do not appear as IPP, even if some of the arguments belonging to the *te*-infinitive appear in the Mittelfeld:

(5.45) a. en dan moet je dat proberen$_1$ alleen maar te doen$_2$ in
and then must.FIN you that try only just to do in
win-winsituaties.
win-win situations
'and then you must only try to do that in win-win situations' [CGN, fvj601241__65]

b. en ik denk dat men daarin moet$_1$ trachten$_2$ het juiste evenwicht te
and I think that one there-in must try the right balance to
zoeken$_3$.
search.INF
'and I think that one has to try to find the right balance in that.' [CGN, fvg600012__38]

Those examples suggest that optionally IPP verbs are not clustering if they do not occur as IPP. The examples in (5.45) are instances of the third construction, as some arguments of the main verb appear in the Mittelfeld while multiple non-verbal elements appear between the verbs. Still, the large amount of constructions that seem to be in between clustering and non-clustering indicates that we need an analysis that adequately accounts for such constructions.

In 81% (158 hits) of the constructions in which the **selector cannot appear as IPP** the verbs are not adjacent, i.e. constructions in which te *te*-infinitive and its arguments are in the Nachfeld, as in (5.46).

(5.46) ... net nu hij dacht$_1$ een goede indruk te maken$_2$.
... just now he thought a good impression to make
'Just as he thought to make a good impression.' [fvo800965__1]

The 37 adjacent examples consist of the third construction (5.47a), constructions in which the *te*-infinitive selects its dependents to the right (5.47b), fixed expressions (5.47c), and some predicative constructions in which the *te*-infinitive is tagged as a VC (5.47d).

(5.47) a. ...omdat hij tijdens de PCB-crisis stalen vergat$_1$ te controleren$_2$.
...because he during the PCB crisis samples forgot to check
'...because he forgot to check samples during the PCB crisis.' [LASSY, WR-P-P-I-0000000165.p.3.s.5]

b. Dit betekent dat u op elk moment kunt$_1$ beslissen$_2$ te verhuizen$_3$ naar
this means that you on each moment can decide to move to
een andere agent.
another agent
'This means that you can decide to change agent at any time' [LASSY, dpc-dns-001069-nl-sen.p.2.s.2]

c. De grote armoede waarin de Weimarrepubliek sinds haar oprichting mee
the great poverty where-in the Weimar Republic since her foundation with
had$_1$ te kampen$_2$ ...
had to fight ...
'The extreme poverty with which the Weimar Republic had to deal from the start ...' [LASSY, WR-P-E-I-0000054957.p.2.s.220.2]

d. ...waar nog het een en ander valt$_1$ te verbeteren$_2$.
...where still it one and other falls to improve
'...where there is still a lot to improve' [LASSY, WR-P-P-L-0000000001.p.98.s.4]

Even though *vallen* 'fall' can select a *te*-infinitive to its right, its passive (modal) meaning, and the direction of selection indicate that it is a predicate selector, rather than a clustering verb.[20]

Summing up, the treebank investigation of verbs that select a *te*-infinitive in the second pole or in the Nachfeld illustrate that the obligatory IPP verbs invariably select their verbal complement in the second pole (also if they do not appear in the perfect tense), rather than in the Nachfeld.

The optional IPP verbs show a more mixed pattern between clustering constructions, the third construction, constructions with an extraposed complement, and ambiguous constructions. If they appear as IPP, they never occur with an extraposed

[20]There are 24 treebank examples in which *vallen* selects a predicative *te*-infinitive; in 21 cases the *te*-infinitive is selected on the left.

te-infinitive, but they can if they do not appear as IPP, cf. (5.43) and (5.45). No constructions were found in which a *te*-infinitive selects another *te*-infinitive in the second pole. Although the literature does not rule out such constructions, it seems that they are avoided in authentic language examples.

The obligatory and optional IPP verbs (374+311) are listed in Table 5.14. The figures present how often they occur as a selector of a *te*-infinitive in the constructions listed in Table 5.12. In addition, it is indicated how many of those constructions are IPP constructions. The clustering constructions thus at least include the 374 cases in which a *te*-infinitive is selected by an obligatory IPP verb and 16 out of the 311 constructions in which a *te*-infinitive is selected by an optional IPP verb that shows the IPP effect, whereas the optional IPP verbs do not (unambiguously) select their complement in the second pole (i.e. the figures in gray).

The verbs in Table 5.14 all select a *te*-infinitive, but several verbs also occur in the set of clustering verbs selecting a participle or a bare infinitive (cf. Table 5.7). As discussed in section 5.2, constructions with *te*-omission are always clustering, but if *te* is present, the optional IPP verbs can also select their complement in the Nachfeld.

The majority of the verbs in Table 5.14 are subject-oriented; *helpen* 'help' is the only object-oriented verb that selects a *te*-infinitive in the treebanks. The set of subject-oriented obligatory clustering verbs includes the modals *hoeven* 'have to', *dienen* 'ought to' and *(be)horen* 'ought to', the evidentials *lijken* 'seem', *blijken* 'turn out', *schijnen* 'appear', and *heten* 'be reputed to', the aspectual verbs and some subject control verbs. The aspectual verbs all appear in Table 5.7 as well, indicating that those verbs typically allow *te*-omission. In contrast to the other aspectual verbs, the meaning of *komen* 'come' in combination with a *te*-infinitive differs from *komen* with a bare infinitive, as is clear from the translations in (5.48a) and (5.19d), repeated as (5.48b).

(5.48) a. Dat is vast komen$_1$ te staan$_2$ na onderzoek door de politie.
that is fixed come to stand after investigation by the police
'That has been found to be the case after investigation by the police.' [LASSY, WS-U-E-A-0000000219.p.3.s.1]

b. want ja je wou er vandaag toch over komen$_1$ praten$_2$?
because yeah you wanted there today actually about come talk
'because you wanted to come and talk about it today, right?' [CGN, fvf600243__110]

	LEMMA	TE INF	IPP
OBLIGATORY IPP VERBS	*(be)hoeven* 'need'	58	3
	dienen 'ought'	34	1
	(be)horen 'ought'	3	0
	lijken 'seem'	39	0
	blijken 'turn out'	26	0
	schijnen 'appear'	6	0
	heten 'be reputed to'	2	0
	zitten 'sit'	53	0
	komen 'come'	41	7
	staan 'stand'	27	0
	liggen 'lie'	10	0
	lopen 'walk'	5	0
	weten 'know (how to)'	47	19
	zien 'manage'	20	0
	plegen 'tend'	3	0
OPTIONAL IPP VERBS	*beginnen* 'begin'	66	4
	proberen 'try'	131	8
	durven 'dare'	19	3
	dreigen 'threaten'	22	0
	weigeren 'refuse'	19	0
	wensen 'wish'	17	0
	trachten 'try'	11	1
	hopen 'hope'	6	0
	beloven 'promise'	4	0
	menen 'mean, intend'	5	0
	leren 'learn'	1	0
	pogen 'try'	1	0
	wagen 'dare'	1	0
	helpen 'help'	8	0
	Sum	**685**	**46**

Table 5.14: Clustering verbs selecting a *te*-infinitive

Subject-oriented *zien* 'manage' differs from object-oriented *zien* 'see' in that the former invariably combines with a *te*-infinitive, whereas the latter never selects a *te*-infinitive. Compare (5.49a) and (5.22c), repeated in (5.49b).

(5.49) a. Bij de beantwoording daarvan zullen we telkens de juiste balans moeten$_1$
by te answering there-of will we each time the right balance must
zien$_2$ te vinden$_3$.
manage to find
By answering that we will have to find the right balance each time.' [LASSY, dpc-bal-001239-nl-sen.p.33.s.3]

b. De vacaturesite Monsterboard.nl heeft het aantal ICT-vacatures in
The job site Monsterboard.nl has the amount ICT vacancies
drie maanden tijd zien$_1$ verdubbelen$_2$.
in three months time see duplicate.
'The job site Monsterboard.nl has seen the amount of ICT vacancies duplicate in three months time.' [LASSY, WS-U-E-A-0000000233.p.14.s.5]

The optionally clustering verbs that select a *te*-infinitive include the aspectual verb *beginnen* 'begin', several subject control verbs (of which *proberen* 'try' is by far the most common one), and object-oriented *helpen* 'help'. As shown in the examples in (5.50), they are always clustering if they appear as IPP (5.50a-b), but in the other constructions they can also select an extraposed complement (5.50c-d).

(5.50) a. ... kort nadat ze voor Helmsley was beginnen te werken in september
... shortly after she for Helmsley was begin to work in September
1983 ...
1983 ...
'... shortly after she had begun to work for Helmsley in September 1983 ...' [LASSY, dpc-ind-001643-nl-sen.p.10.s.2]

b. ... heb 'k ik daar ene proberen te bellen ...
... have I I there one try to call ...
'... I have tried to call one there ...' [CGN, fvd700066__189]

c. ... dat hij in de toekomst gaat proberen kinderen met psychische
... that he in the future goes try children with psychological
problemen niet meer bij criminelen op te sluiten.
problems not anymore with criminals up to lock.INF
... that he will try not to lock up children with psychological problems together with other criminals in the future. [LASSY, WS-U-E-A-0000000040.p.21.s.4]

d. Het ... team ... moet de Afghaanse regering helpen het gezag in de
The ... team ... must the Afghan government help the power in the
noordelijke provincie te herstellen.
northern province to restore
'The ... team ... needs to help the Afghan government to re-establish its power in the northern province.' [LASSY, WS-U-E-A-0000000233.p.12.s.2]

As several of the verbs in Table 5.14 do not appear as IPP, empirical evidence for their IPP status was found on the Web, cf. section 5.4.

5.3.4 Conclusion

In contrast to verbs selecting a past participle or a bare infinitival complement, the constructions with a *te*-infinitive encountered in the treebanks show a more mixed picture between verb clusters and non-clustering constructions, i.e. constructions with a verbal complement in the Nachfeld, predicative constructions, and fixed expressions.

Whereas verbs selecting a past participle or a bare infinitive are typically clustering, verbs selecting a *te*-infinitive may be clustering or non-clustering. The treebank examples indicate that obligatory IPP verbs are always clustering, while the optionally IPP verbs can also select an extraposed complement if they do not occur as IPP. If a *te*-infinitive occurs in front of its selector, it is typically a predicative complement or a part of a fixed expression, rather than the complement of a clustering verb.

The typology in Figure 5.3 provides an overview of the different verb categories that unambiguously select a *te*-infinitive in the cluster, as well as their relative frequency in the treebanks, based on Table 5.14.

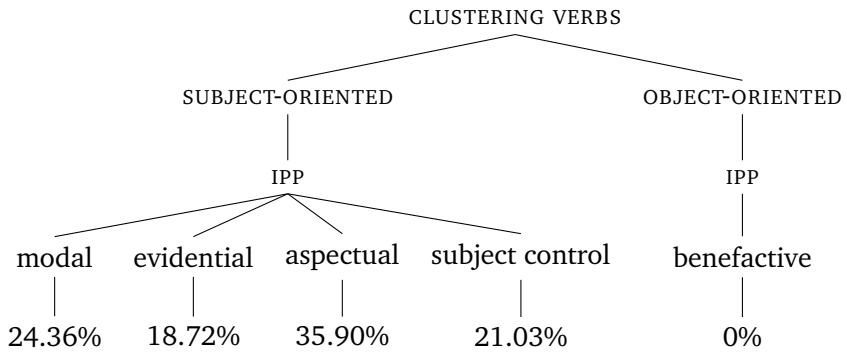


Figure 5.3: Treebank-based typology of clustering verbs selecting an *te*-infinitive

As there are no benefactives selecting a *te*-infinitive that appear as IPP in the data, there are no treebank instances in which they unambiguously select their complement in the second pole. Hence, the frequency in Figure 5.3 is 0% for that category.

5.4 IPP constructions

As motivated in the previous sections, IPP is an important diagnostic for clustering verbs. In section 5.2 it was mentioned that the majority of the verbs selecting a past participle and/or a bare infinitive are obligatory or optional IPP verbs. In section 5.3 it was shown that obligatory IPP verbs are always clustering, whereas optional IPP are invariably clustering if they appear as IPP.

Although there are many lists of IPP verbs available in the literature (see chapter 1), several authors disagree about the occurrence of IPP for some verbs and about the optionality of the phenomenon. Moreover, none of the authors cited in chapter 1 claim to provide an exhaustive list, even though the set of IPP verbs is assumed to be a limited set of verbs.

In this section it will be shown how IPP constructions were extracted from the treebanks (section 5.4.1). Next, it will be investigated which verbs obligatorily appear as IPP and which verbs optionally occur as IPP in the treebanks (section 5.4.2). As several IPP verbs mentioned in the literature do not occur as IPP in the treebanks due to data sparseness, evidence for their IPP status was found on the Internet using Google search. Section 5.4.3 concludes.

A first treebank-based investigation of Dutch IPP verbs is described in Augustinus & Van Eynde (2012).[21] A comparative typology of Dutch and German IPP verbs is presented in Augustinus & Van Eynde (in press).

5.4.1 Extracting IPP constructions

The IPP verbs in the treebanks were retrieved by means of GrETEL, as described in section 4.3.3. The automatically generated query in (4.11) was manually adapted to (4.14), repeated in (5.51), in order to include all IPP constructions in the treebanks.

(5.51)
```
//node[@cat and node[@rel="hd" and @pt="ww" and
(@lemma="hebben" or @lemma="zijn")] and node[@rel="vc" and
@cat="inf" and node[@rel="hd" and @pt="ww"] and node[@rel="vc"
and (@cat="inf" or @cat="ti" or @pt="ww")]]]
```

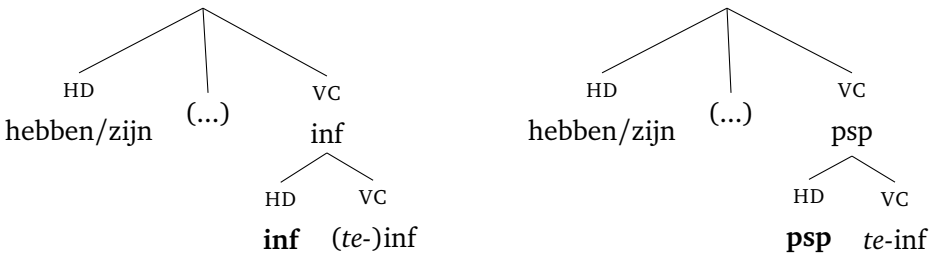
[21] Some of the treebank counts presented in Augustinus & Van Eynde (2012) may slightly differ from the counts presented here, as a previous version of the LASSY treebank was used for that article. Moreover, the (Alpino) VERB tag was used instead of the WW tag to retrieve verbs, which sometimes returns slightly different results. In the results presented here, the same queries were used for both the CGN and LASSY treebanks.

Besides IPP constructions, the corresponding constructions with a past participle were extracted as well, in order to investigate the verbs that optionally appear as IPP. Those constructions, in which a verb of the perfect selects a past participle followed by a *te*-infinitive, are henceforth referred to as *PSP constructions*. Extracting PSP constructions was done using the query in (5.52) which is very similar to the query for IPP constructions in (5.51): Only the phrasal tag of the first VC node was changed from infinitival to participial (`@cat="ppart"`), and the form of the infinitival complement is always a *te*-infinitive, as past participles cannot take a bare infinitival complement in Dutch.

(5.52)
```
//node[node[@rel="hd" and @pt="ww" and (@lemma="hebben"
or @lemma="zijn")] and node[@rel="vc" and @cat="ppart" and
node[@rel="hd" and @pt="ww"] and node[@rel="vc" and (@cat="ti"
or @pt="ww")]]]
```

A graphical representation of the queries in (5.51) and (5.52) is presented in (5.53).

(5.53)



Note that in verb-initial clauses the finite verb and the other verbs are not necessarily adjacent. This is indicated in (5.53) by (...). As XPath performs a *greedy* search, it is not necessary to specify this in the XPath expressions (cf. section 4.3.3). Due to some idiosyncratic treebank annotations, some complements of IPP verbs are not tagged as a VC but as a separable verb particle (SVP).[22]

[22]For example *weten* 'know' in *Dat heeft hij tijdens een persconferentie laten weten.* 'He has announced that during a press conference'. [LASSY, WS-U-E-A-0000000205.p.22.s.2]. Those constructions were extracted using the query `//node[node[@rel="hd" and @pt="ww" and (@lemma="hebben" or @lemma="zijn")] and node[@rel="vc" and @cat="inf" and node[@rel="hd" and @pt="ww"] and node[@rel="svp" and @pt="ww"]]]`.

5.4.2 A typology of IPP verbs

Table 5.15 presents the general counts for both IPP constructions and PSP constructions, after manually filtering out erroneous results.[23] In the set of PSP constructions, the instances with infinitival predicative complements were excluded from the data set.[24]

	CGN	LASSY	**Sum**
IPP	792	309	1,101
PSP	31	125	156
Sum	**823**	**434**	**1,257**

Table 5.15: IPP and PSP constructions in CGN and LASSY Small

The IPP constructions are a small subset of the constructions with verb clusters discussed in sections 5.2 and 5.3. The results in Table 5.15 reveal that in both treebanks IPP constructions occur more often than PSP constructions. Having extracted the relevant constructions, it is possible to collect information about the individual IPP verbs. Table 5.16 provides a quantitative survey of the IPP verbs in the CGN treebank and LASSY Small. The data of both treebanks are merged, as for some verbs the number of IPP constructions is small. For each verb Table 5.16 specifies the number of hits for the IPP form in combination with a bare infinitive (IPP + INF), the IPP form in combination with a *te*-infinitive (IPP + TE INF), and the PSP form in combination with a *te*-infinitive (PSP + TE INF).

The numbers for IPP add up to 1,101, which is the total number of IPP constructions in the treebanks, while the numbers for PSP constructions add up to 52, which is a subset of the PSP constructions in Table 5.15, as the constructions with a past participle that never appear in the IPP form (e.g. *vragen* 'ask') are not included in Table 5.16. Besides information about IPP, Table 5.16 provides information about the adjacency of the verbs and their infinitival complement, i.e. ADJ(ACENT) versus NOT ADJ(ACENT).

[23] Sentences containing annotation errors or disfluencies were excluded from the data set, such as *ja nogal wiedes ja want je hebt een tijdje daar in die diepvrieskist daar zitten ...* [CGN, fna000724__19], which lacks an explicit second infinitive but has a tag for it.

[24] For instance *Dat heeft het onderzoeksteam te horen gekregen.* 'The research team got to hear that.' [LASSY, WS-U-E-A-0000000015.p.11.s.2].

		Lemma	IPP + INF		IPP + TE INF		PSP + TE INF	
			ADJ	NOT ADJ	ADJ	NOT ADJ	ADJ	NOT ADJ
SUBJECT-ORIENTED VERBS	OBLIGATORY IPP	*kunnen* 'can'	209	4	–	–	–	–
		moeten 'must, have to'	186	5	–	–	–	–
		mogen 'may'	28	1	–	–	–	–
		zullen 'shall, will'	1	0	–	–	–	–
		hoeven 'need'	4	0	2	1	–	–
		dienen 'ought'	0	0	1	0	–	–
		gaan 'go, will'	185	3	–	–	–	–
		blijven 'remain, continue'	28	1	–	–	–	–
		wezen 'be in the process of'	13	0	–	–	–	–
		komen 'come'	20	0	7	0	–	–
		zitten 'sit'	44	0	0	0	–	–
		staan 'stand'	15	0	0	0	–	–
		lopen 'walk'	5	0	0	0	–	–
		willen 'want'	53	0	–	–	–	–
		weten 'know, manage'	0	0	16	3	–	–
	OPTIONAL IPP	*beginnen* 'start, begin'	7	1	4	0	2	5
		leren 'learn'	19	0	0	0	0	1
		durven 'dare'	5	0	2	1	0	0
		proberen 'try'	1	0	6	2	4	22
		trachten 'try'	0	0	0	1	0	1
		pogen 'try'	0	0	0	0	0	1
		weigeren 'refuse'	0	0	0	0	0	1
		dreigen 'threaten'	0	0	0	0	0	6
		beloven 'promise'	0	0	0	0	0	9
OBJECT-ORIENTED VERBS	OBLIGATORY IPP	*laten* 'let'	153	1	–	–	–	–
		doen 'do'	6	0	–	–	–	–
		zien 'see'	36	0	–	–	–	–
		horen 'hear'	19	0	–	–	–	–
	OPTIONAL IPP	*leren* 'teach'	1	0	0	0	0	0
		helpen 'help'	1	0	0	0	0	0
Sum			**1,039**	**16**	**38**	**8**	**6**	**46**

Table 5.16: Treebank-based typology of IPP verbs

The IPP verbs are divided according to the same mutually independent distinctions used for the classification of the clustering verbs selecting a *te*-infinitive (Table 5.14), i.e. they are split up into subject-oriented and object-oriented verbs on the one hand and into obligatory IPP and optional IPP verbs on the other hand. If a verb cannot occur in a certain construction, e.g. if it cannot appear in the PSP form in combination with an auxiliary of the perfect, it has '–' in that column. If a verb can occur in a certain construction but if no treebank instances were found (due to data sparseness), it has a '0' in that column.

IPP and adjacency

The distinction between adjacent and non-adjacent constructions shows a marked contrast between the IPP forms and the PSP constructions: While the split cases account for 24 of the 1,101 IPP constructions (i.e. a meagre 2.18 %), they account for 46 of the 52 combinations with PSPs (i.e. 88.46%).

The non-adjacent IPP constructions are instances with cluster creepers, such as separable particles, stranded adpositions, and idiom chunks (5.54a), while the non-adjacent constructions with past participles tend to be full VPs (5.54b). This contrast nicely shows that infinitives that are selected by an IPP are part of the verb cluster, while infinitives selected by a past participle are in the Nachfeld.

(5.54) a. ze had het er nog niet kunnen tegen zeggen.
she had it there still not can.IPP to say
'She hadn't been able to say it to him.' [CGN, fva400284__24]

b. Amerika heeft beloofd gelijksoortige maatregelen te nemen.
America has promised similar measures to take
'The US has promised to take similar measures.' [LASSY, WR-P-P-I-0000000242.p.5.s.7]

In (5.54a) the modal *kunnen* 'can' is separated from its complement by a stranded adposition. A more detailed discussion of instances of cluster creeping is presented in section 5.5. In (5.54b) the past participle *beloofd* 'promised' selects the *te*-infinitive in the Nachfeld, from which it is separated by the non-verbal complement *gelijksoortige maatregelen*.

Subject-oriented IPP verbs

The subject-oriented obligatory IPP verbs include the modals, two types of aspectual verbs and some subject control verbs. The subject-oriented optional IPP verbs include the aspectual verb *beginnen* 'begin' and a set of subject control verbs.

The **modals** comprise the core modals *kunnen* 'can', *moeten* 'must', *mogen* 'may' and *zullen* 'shall', as well as *hoeven (te)* 'need (to)' and *dienen te* 'ought to'. They jointly account for 442 of the 1,101 IPPs, which amounts to 40.14 %. They all take *hebben* 'have' as the auxiliary of the perfect (5.55a). However, if the selected infinitive requires *zijn* 'be', the latter may be used instead. This is attested for *moeten* (7 hits), *kunnen* (3 hits) and *mogen* (1 hit). An examples is given in (5.55b). This phenomenon is described in amongst others Draye & van der Horst (2006); a treebank investigation is conducted in Van Eynde et al. (in press).

(5.55) a. Pas nu hebben we dat ook kunnen zien in de hersenen.
only now have we that also can.IPP see in the brains
'Only now we have been able to see that in the brains.' [LASSY, dpc-ind-001634-nl-sen.p.16.s.5]

b. . . . dat er . . . een gelegenheidsstructuur is kunnen ontstaan die allerlei
. . . that there . . . a ad hoc structure is can.IPP emerge that all sorts
criminele activiteiten faciliteert.
criminal activities facilitates
'. . . that an ad hoc structure has been able to emerge that facilitates all sorts of criminal activity.' [LASSY, dpc-bal-001237-nl-sen.p.8.s.5]

The **aspectual** obligatory IPPs come in two types. The first type includes *blijven* 'stay, keep, remain', *gaan* 'go', *wezen* 'be', and *komen* 'come'. They take a form of *zijn* as the auxiliary of the perfect, but if the selected infinitive requires a form of *hebben*, the latter may be used instead. This is attested for *gaan* (4 hits), *komen* (3 hits) and *wezen* (2 hits). Some examples with *komen* are given in (5.56).

(5.56) a. en Erwin Jans die is er weer bij komen zitten want . . .
and Erwin Jans who is there again with come.IPP sit because . . .
'and Erwin Jans, he has come to join us because . . .' [CGN, fvl600281__1]

b. heeft er niemand dat komen zeggen tegen jullie?
has there no one that come.IPP say to you
'Did nobody come to tell you that?' [CGN, fva400386__18]

c. Dat is vast komen te staan na onderzoek door de politie.
that is fixed come.IPP to stand after investigation by the police
'That has been found to be the case after investigation by the police.' [LASSY, WS-U-E-A-0000000219.p.3.s.1]

Komen can combine with bare infinitives and *te*-infinitives, but with a different meaning, as is clear from the translations.

The second type of aspectual IPPs include the posture verbs *zitten* 'sit', *staan* 'stand' and the motion verb *lopen* 'run'. They typically select a *te*-infinitive, but they allow *te*-omission, especially in IPP contexts, such as (5.57). In contrast to the former type, they invariably select *hebben* 'have' as the perfect auxiliary.

(5.57) heb ik uh een hele lange brief zitten schrijven.
have I uh a very long letter sit.IPP write
'I have uh been writing a very long letter.' [CGN, fna000600__43]

The aspectual obligatory IPP verbs jointly account for 321 of the 1,101 IPPs, which amounts to 29.15%.

Completing the survey are the **subject control** verbs *willen* 'want' and *weten te* 'manage to'. The former takes a bare infinitive and the latter a *te*-infinitive. Within the class of subject-oriented obligatory IPPs, the subject control verbs are the smallest group: There are only two of them and they jointly account for 72 of the 1,101 IPPs, which amounts to a modest 6.54%.

The optionally subject-oriented IPP verbs include the **aspectual** *beginnen* 'begin' and a number of **subject control** verbs. They all select a *te*-infinitive, as shown for trachten 'attempt' in (5.58a), but some of them allow *te*-omission, as shown for *leren* 'learn' in (5.58b).

(5.58) a. ...en we hebben trachten duidelijk te maken dat ...
...and we have try.IPP clear to make that ...
'...and we have tried to make it clear that ...' [LASSY, dpc-cam-001017-nl-sen.p.14.s.2]

b. Hij heeft leren rekenen in ezels, kippen en schapen.
he has learn.IPP count in donkeys chickens and sheep
'He has learned to count in donkeys, chickens and sheep.' [CGN, fvj601101__27]

Many of the verbs in this category only occur in the PSP form in the treebanks, but this is due to data sparseness. Examples of their IPP forms can be found on the Web, as illustrated for *dreigen* 'threaten' and *beloven* 'promise' in (5.59).[25]

(5.59) a. Die ene kus had geleid tot een volgende en was ontvlamd in een vuurzee
that one kiss had led to a next and was sparked in a see of fire
die hen allebei had dreigen te overspoelen.
that them both had threaten.IPP to overwhelm
'That one kiss had led to another one and had sparked a sea of flames that had threatened to overwhelm them both.' [Google.be, 13-01-2015]

b. Ik vraag me af welke film Will heeft beloven te doen, zodat hij die rol
I ask me off which film Will has promise.IPP to do, so he that role
kreeg.
get
'I wonder which movie Will has promised to do in order to get that role.' [Google .be, 13-01-2015]

Conversely, *durven* 'dare' only occurs in the IPP form in the treebanks, but instances of its PSP form can be found on the Web, as in (5.60).

(5.60) Men heeft gedood, verwond en zonder onderscheid gevangen genomen wie
one has killed wounded and without distinction imprisoned taken who
terecht heeft gedurfd te protesteren.
rightly has dared to protest
'one has killed, wounded and imprisoned indiscriminately whoever dared to protest' [Google.nl, 13-01-2015]

The subject control verbs of this class all take *hebben* as the auxiliary of the perfect. Combinations with *zijn* are not attested in the treebanks, but can be found on the Web, at least for *durven*.

(5.61) Ik vind het enorm moedig dat ze tot bij mij is durven komen.
i find it very courageous that she to with me is dare.IPP come
'I consider it very courageous that she has dared to come to me.' [Google.be, 13-01-2015]

Although the number of subject-oriented optional IPPs is relatively large, their frequency in the treebanks is small: They jointly account for 49 of the 1,101 IPPs, which amounts to a meagre 4.45%.

[25] Similar examples can be found for *pogen* 'attempt' and *weigeren* 'refuse'.

Object-oriented IPP verbs

The object-oriented obligatory IPP verbs include the **causatives** *laten* 'let' and *doen* 'do', and the **perception verbs** *zien* 'see' and *horen* 'hear', as used in (5.62).

(5.62) Ik heb 't twee keer zien gebeuren.
I have it two times see.IPP happen
'I have seen it happen twice.' [CGN, fna000773__212]

They take *hebben* as the auxiliary of the perfect and invariably select a bare infinitive, which implies that they lack a PSP counterpart. The causatives account for 160 hits (14.53%) and the perception verbs for another 55 (5%).

The object-oriented optional IPP verbs comprise the **benefactives** *helpen* 'help' and *leren* in the sense of 'teach'. They select a *te*-infinitive, but allow *te*-omission. The treebanks contain only two hits for this class (0.18%). They are presented in (5.63).

(5.63) a. 'k heb mijn kleine kinderen daar ook leren zwemmen.
I have my little children there also teach.IPP swim
'I have also taught my little children how to swim over there.' [CGN, fva400659__44]

b. Dit onderzoek heeft ... de modaliteiten van enkele later geïmplementeerde
this research has ... the terms of some later implemented
beleidsmaatregelen mee helpen bepalen.
policy measures with help.IPP define
'This research has helped to define the terms of some policy measures that were implemented later.' [LASSY, dpc-fsz-000551-nl-sen.p.22.s.4]

There are no hits for their PSP counterparts, but examples can be found on the Web, as shown for *helpen* 'help' in (5.64).

(5.64) Deze reactie is een oeroud mechanisme dat ons heeft geholpen te overleven
this reaction is a very old mechanism that us has helped to survive.INF
tijdens acuut gevaar
during critical danger
'This reaction is an ancient mechanism that has helped us to survive during critical danger.' [Google.nl, 14-08-2014]

Unattested IPPs

This concludes the survey of the Dutch IPP verbs that occur in the treebanks. Some of them only show up in their PSP form, but are included in Table 5.16, since their IPP

counterparts can be found on the Web. In spite of their inclusion, the classification cannot be considered complete. This is due to the fact that several of the IPP verbs are so uncommon that they do not occur in the treebanks, neither in their IPP form nor in their PSP form. This is for instance the case for the evidential modals *blijken* 'turn out', *lijken* 'appear', *schijnen* 'seem', and *heten* 'be reputed to', as shown in (5.65). Since their PSP counterpart cannot be used instead, they are obligatory IPP verbs, confirming their treatment in Rutten (1991).

(5.65) a. Mogelijk omdat licht in die rustige 20e eeuw minder belangrijk heeft
possible because light in that quiet 20th century less important has
schijnen te zijn.
seem.IPP to be
'Possibly because light has seemed to be less important in that quiet 20th century.' [Google.be, 17-07-2014]

b. Patricia, die altijd het lievelingetje had heten te zijn, werd nu
Patricia, who always the favourite.DIM had be called.IPP to be was now
genegeerd.
neglected
'Patricia, who always was reputed to be the favourite one, was now neglected.' [Google.be, 13-01-2015]

Other candidates for inclusion in the class of the subject-oriented obligatory IPP verbs are the modal *horen* 'ought', the aspectual *liggen* 'lie', and the subject control verbs *zoeken* 'intend', and *zien* 'manage'. An example is given in (5.66).

(5.66) ... als hij niet voldoende feiten heeft zien te vinden om het bestaan en
... if he not enough facts has manage.IPP to find to the existence and
de hoogte van het belastbaar inkomen te bepalen.
the height of the taxable income to determine
'... if he didn't manage to find enough facts to determine the existence and the amount of the taxable income' [Google.be, 14-08-2014]

Notice that the subject-oriented *horen* 'ought' and *zien* 'manage' require a *te*-infinitive, while the homophonous perception verbs require a bare infinitive.

The verbs *durven* 'dare', *beloven* 'promise', and *dreigen* 'threaten' are included in Table 5.16 as optional IPP verbs, but if they are used in the sense 'be on the point of', they obligatorily appear as IPP in the perfect tense. In that sense, they also allow non-referential subjects and (inanimate) non-agentive subjects, in contrast to the literal variants listed in Table 5.16. An example with *dreigen* was given in (5.59a).

In the same vein, the class of subject-oriented optional IPPs can be extended with the subject control verbs *begeren* 'desire', *hopen* 'hope to', *wensen* 'wish', *menen* 'mean', *wagen* 'risk' and *vermogen* 'be able', as used in (5.67).

(5.67) Zit daar een diepere bedoeling achter, die ik niet heb vermogen te
sit there a deeper intention behind that I not have be able.IPP to
doorgronden?
see through
'Is there an underlying intention, that I haven't been able to see through?' [Google .be, 17-07-2014]

An extension for the class of object-oriented obligatory IPPs is the perception verb *voelen* 'feel', as used in (5.68).

(5.68) Hij vertelt het zonder emotie, alsof hij dit al lang heeft voelen aankomen.
he tells it without emotion as if he this already long has feel.IPP arrive
'He is telling it without emotion, as if he has felt it coming for a long time.' [Google .nl, 14-08-2014]

Summing up, the Dutch IPP verbs form a relatively large and heterogeneous group. Those that are attested in the treebanks can be partitioned in terms of two mutually independent distinctions, i.e. subject-oriented versus object-oriented, and obligatory versus optional. Unattested members can be fitted into one of these four classes.

5.4.3 Conclusion

This section illustrated how IPP constructions can be extracted from the treebanks and how the data support a classification into subject-oriented versus object-oriented IPP verbs, on the one hand, and obligatory versus optional IPP verbs, on the other hand. The data do not contain any IPP verbs that are not listed in the literature on the topic, and due to data sparseness, not all verbs encountered in the literature occur as IPP in the treebank data. For those verbs, their IPP status was defined by means of Internet data, in order to construct a typology that is based on empirical data.

The classification presented here only contains **Dutch IPP verbs**, but in Augustinus & Van Eynde (in press) it is shown that it is also suitable to describe the **German IPP verbs**, and that the classification is more accurate than existing typologies, such as the one proposed in Schmid (2005).

In addition to the typology, the data illustrate that IPP constructions are clustering constructions. The majority of the IPP verbs are adjacent to the infinitival complement they select, in contrast to the infinitives selected by a past participle. Another characteristic indicating the clustering properties of IPP verbs is *te*-omission, which occurs in a considerable number of constructions. As a bare infinitive is never selected in the Nachfeld, such constructions are invariably clustering.

5.5 Cluster creepers

While the previous sections focussed on the verbs constituting a verb cluster, this section draws attention to the non-verbal elements that may appear in clusters. The treebank investigation presented here not only provides insight in the frequency of the phenomenon; it also results in a classification of the types of cluster creepers according to their syntactic function and their lexical (POS) or phrasal category. This makes it possible to investigate whether the types of cluster creepers mentioned in Haeseryn et al. (1997) (see section 1.6) are reflected in the corpus data, or whether the data reveal other categories, aiming at a more complete description of the possible cluster creepers in Dutch.

Furthermore, the occurrence of cluster creepers in spoken versus written language will be discussed, as well as cluster creeping in clusters containing participles versus clusters with infinitives. Some relevant examples are given in (5.69).

(5.69) a. we hebben zo nog ne politieker die ons daar altijd ook doet *aan*
we have so another a politician that us there always also does on
denken.
think
'we have another politician of that kind who always reminds us of that. [CGN, fvc701156__222]'

b. . . . aan iedereen die toen de toekomst van dit land, van de huidige en
. . . to everyone that then the future of this country of the present and
toekomstige generaties hebben *veilig* gesteld.
future generations has safe put
'. . . to everyone who back then has safeguarded the future of this country, of the current and future generations.' [LASSY, dpc-vhs-000745-nl-sen.p.13.s.3]

Section 5.5.1 explains how constructions containing cluster creepers were extracted from the treebanks. Section 5.5.2 presents and discusses the results of the

treebank investigation. Those results largely confirm the claims made in Haeseryn et al. (1997), but they also contain a surprise, i.e. *multiple cluster creepers*, as in (5.70).

(5.70) ... dat we ons daar nog kunnen *mee bezig* houden.
... that we us there still can with busy keep
'... that we can still keep ourselves busy with that.'

Section 5.5.3 sums up the conclusions and points out some topics for future research. An earlier investigation on cluster creeping in the CGN and LASSY treebanks is described in Augustinus & Van Eynde (2014).[26]

5.5.1 Extracting constructions with cluster creepers

After having collected the set of verb clusters by means of XPath queries as described in sections 5.2.1 and 5.3.1, the constructions containing cluster creepers can be extracted, i.e. constructions in which non-verbal elements occur between the verbs in the second pole. Since it is hard to determine the linear order of the nodes in an elegant way using XPath, XQuery scripts were used in order to identify the constructions in which non-verbal elements occur between the verbs by means of word order constraints. As an example, the XQuery script used to find cluster creepers in two-verb clusters with a finite verb and an infinitive is included in appendix C.

Due to the treebank design, it is not possible to extract all constructions with cluster creepers in that way, however. Separable verb particles (SVPs) are only tagged separately if they are written as a separate word (e.g. *zal koffie drinken* 'will coffee drink'), but not if they are written as a part of the verb (e.g. *zal koffiedrinken*). In LASSY Small they can be extracted in another way, but not in the CGN treebank, as will be explained in section 5.5.2. Therefore, the focus will be on the extraction of cluster creepers that are written as a separate word. Since the set of constructions with cluster creepers is low in comparison to the set of all clustering constructions, the results were manually verified after the automatic extraction.

Even though the quality of the annotations in both LASSY and CGN is very high, the treebanks contain some annotation errors that are problematic for this case study. For example, sentences that are erroneously tagged as verb-final whereas they are verb-initial.

[26] With respect to clusters containing a *te*-infinitive, the treebank investigation in Augustinus & Van Eynde (2014) only includes IPP constructions.

In (5.71), for instance, the clause after 'uh' is tagged `ssub` instead of `smain`.

(5.71) dan kan ik uh ik kan 'm in de keuken nergens inpluggen vrienden.
then can I uh I can him in the kitchen nowhere plug in friends
'then I can't plug it in in the kitchen, friends.' [CGN, fna000573__58]

Besides the elimination of annotation errors, two types of false positives were filtered out semi-manually. The first type concerns constructions with stopgaps, corrections, and/or interruptions, such as the examples in (5.72). Those constructions were mainly encountered in CGN.

(5.72) a. maar wat wij merkten in Frankrijk was dikwijls dat ge 's middags
but what we noticed in France was often that you at lunchtime
soms zeer goede menu's kondt gebr- allee eten dus hè.
sometimes very good menus could use.INTERRUPTED go.STOPGAP eat thus hè
'What we often noticed in France was that you sometimes could use- well eat very good menus at lunchtime.' [CGN, fva400295__400]

b. enfin ik weet niet hoe ik het moet uh omschrijven uh.
well I know not how I it must uh describe uh
'well I don't know how I have to uh describe it.' [CGN, fva400534__85]

The second type of false positives is the occurrence of punctuation marks within the verb cluster. Those examples were exclusively found in LASSY.

(5.73) Het is dus niet zo dat deze tanks al eerder "gekannibaliseerd" waren om
it is thus not so that this tanks already before "cannibalized" were for
er bruikbare onderdelen uit te halen.
there usable parts out to get
'It is thus not the case that these tanks were "cannibalized" before to get useful parts out of it.' [LASSY, WR-P-E-I-0000013937.p.4.s.235]

5.5.2 Results

Table 5.17 presents the constructions with verb clusters that are unambiguously clustering in the treebanks. They include the constructions without a *te*-infinitive discussed in section 5.2, the constructions in which the selector of the *te*-infinitive is in the first pole and in which the *te*-infinitive selects a participial or a bare infinitival complement (e.g *(schijnt) . . . te kunnen komen* '(seems . . .) to be able to come'), the constructions in which a *te*-infinitive is selected by an obligatory IPP verb in the second pole, and the constructions in which a *te*-infinitive is selected by an optional IPP verb in the second pole in which the IPP effect shows up.

Since cluster creeping is excluded in constructions in which the main verb occurs at the beginning of the cluster, the results were split up into two categories: Clusters in which the main verb is not the first verb in the cluster (MV $\neq$ 1), and clusters in which it is the first verb (MV = 1).

	CGN		LASSY		**Sum**
	MV $\neq$ 1	MV = 1	MV $\neq$ 1	MV = 1	
PARTICIPIAL	1,980	2,168	4,975	2,408	11,531
INFINITIVAL	5,906	45	4,190	6	10,147
Sum	**7,886**	**2,213**	**9,165**	**2,414**	**21,678**

Table 5.17: Constructions with verb clusters in CGN and LASSY

The results show that the proportion of clusters that potentially contain cluster creepers, i.e. the clusters in which the main verb is not the first verb in the cluster (MV $\neq$ 1), is more or less equal in both treebanks, i.e. 78.09% in CGN and 79.15% in LASSY. In the clusters containing a past participle there is a large set in which the main verb (i.e. the past participle) is the first verb in the cluster, whereas this is hardly the case in infinitival clusters.

Table 5.18 presents the results for cluster creeping in both treebanks. Compared to the large amount of clustering constructions, the results in Table 5.18 show that cluster creeping is a very infrequent phenomenon in both CGN and LASSY. In the CGN treebank, 194 constructions with cluster creepers were found, whereas in LASSY only 38 were encountered.

	CGN	LASSY	**Sum**
PARTICIPIAL	27	12	39
INFINITIVAL	167	26	193
Sum	**194**	**38**	**232**

Table 5.18: Frequency of cluster creepers in CGN and LASSY

As expected, cluster creeping occurs more frequently in the spoken data (CGN) than in the written data (LASSY). The constructions account for 1.92% of all clusters that potentially contain cluster creepers (MV $\neq$ 1) in CGN, and for less than 0.33% of those constructions in LASSY. The data furthermore confirm that cluster creeping occurs more often in infinitival clusters than in clusters with a participle.

Single cluster creepers

Despite the low number of corpus examples, the constructions with cluster creepers show a large variety with respect to the type of cluster creepers, both in category and syntactic function. The three types mentioned in Haeseryn et al. (1997) are all present in the data: The sentences in (5.74) show cluster creeping by a predicative adjective (5.74a) and by a part of a fixed expression (5.74b). (5.75) is an example of adposition stranding within the cluster. In (5.76a) the cluster is interrupted by an object, and in (5.76b) by an adverbial modifier.

(5.74) a. de dokters zeggen wel dat 't gaat *goed* komen.
the doctors say that it goes good come
'The doctors say that it will be fine.' [CGN, fva400370__6]

b. 'k zeg dat gaat moeten beginnen *op gang* komen hè.
I say that goes must begin on pace come
'I say that should start to get going.' [CGN, fva400643__87]

(5.75) De plicht die hem nu roept, kan hem straks de mooiste baan kosten
the duty than him now calls can him later the most-beautiful job cost
waar een Beier kan *van* dromen.
where a Bavarian can of dream
'The duty that calls him now can cost him the most beautiful job a Bavarian can dream of.' [LASSY, WR-P-P-I-0000000033.p.21.s.4]

(5.76) a. als ze moeten *teksten* schrijven dan schrijven ze die met de PC.
if they must texts write then write they them with the PC
'If they have to write texts then they write them with a PC.' [CGN, fvb400165__130]

b. maar normaal moet ge dat kunnen *zo* regelen dus dat dat wegblijft
but normally must you that can so arrange thus that that away-stays
dus dat 't niet verschijnt.
thus that it not appears
'But normally you have to arrange that in such a way that it stays away so that it does not appear.' [CGN, fva400079__264]

An overview of all creeper types is provided in Table 5.19. The labels in the columns indicate the syntactic function (dependency relation): Separable verb particle (SVP), prepositional complement (PC), direct object (OBJ1), predicative complement (PREDC), location or direction complement (LD), indirect object (OBJ2), modifier (MOD), and predicative modifier (PREDM). The left part of the table concerns complements selected by the verb, whereas the right part concerns modifiers.

The labels in the rows indicate the lexical categories (POS) at the top half of the table and the phrasal categories at the bottom part of the table. 14 instances of cluster creeping show a combination of several categories. They are not included in Table 5.19, but will be discussed in this section as well.

	SVP	PC	OBJ1	PREDC	LD	OBJ2	MOD	PREDM	SUM (#)	SUM (%)
PREP	31	37	0	2	12	0	7	0	**89**	**40.83**
ADJ	16	0	0	20	0	0	11	0	**47**	**21.56**
N	6	0	17	0	0	0	0	0	**23**	**10.55**
ADV	6	0	0	0	2	0	6	1	**15**	**6.88**
PRON	0	0	4	1	1	0	5	0	**11**	**5.05**
PP	4	1	0	2	7	1	5	0	**20**	**9.17**
NP	0	0	8	0	0	0	1	0	**9**	**4.13**
AP	0	0	0	1	0	0	2	0	**3**	**1.38**
ADVP	0	0	0	0	0	0	1	0	**1**	**0.46**
SUM (#)	**63**	**38**	**29**	**26**	**22**	**1**	**38**	**1**	**218**	
SUM (%)	**28.90**	**17.43**	**13.30**	**11.93**	**10.09**	**0.46**	**17.43**	**0.46**		**100**

Table 5.19: Types of cluster creepers in CGN and LASSY

As expected, the largest category consists of cluster creepers where an SVP occurs within the cluster, as in (5.77).

(5.77) Ik heb mijn agenda niet hoeven *om* te gooien om die man te kunnen
I have my agenda not need over to throw to that man to can
ontvangen ...
receive
I did not have to completely change my schedule to be able to see that man ...'
[LASSY, dpc-rou-000479-nl-sen.p.10.s.14]

As mentioned above, the results do not include the cases of cluster creeping with separable verbs in which the SVP and the verb are written as one word.

Another major group are the prepositional complements. They include the cases of adposition stranding illustrated in (5.75). The other frequently occurring creeper types are also mentioned in Haeseryn et al. (1997), i.e. predicative adjectives (5.74a), direct objects (5.76a), and modifiers (5.76b).

More remarkable examples in the data set are the constructions in which a full phrase occurs within the cluster, such as the prepositional indirect object in (5.78a) and the location complement in (5.78b).

(5.78) a. . . . 'k weet 'k ik niet of dat 'k ik dat nu moet laten weten aan hem of dat 'k
. . . I know I I not or that I I that now must let know to him or that I
ik dat eerst moet *aan mijn kot* vragen . . .
I that first must to my student's apartment ask . . .
'. . . I don't know whether I should let him know or that I should ask (the people of) my student's apartment first.' [CGN, fva400507__4]

b. dat die nu moet *in de Verenigde Staten* blijven in Miami bij de familie . . .
that that now must in the United States stay in Miami with the family . . .
'that he now has to stay in the United States in Miami with his family . . .' [CGN, fvj600261__9]

The four instances of phrasal SVPs in Table 5.19 all contain fixed expressions, such as the example given in (5.74b). Non-verbal parts of fixed expressions are tagged as SVPS in the treebanks, but one could also classify those constructions as PCS.

The treebank data can be compared to the hierarchy of cluster creepers proposed by Wurmbrand (2005), which was discussed in section 1.6 and repeated in Table 5.20.

Separable particles	Adverbs Idioms Bare nouns	Indefinite objects PPs	Definite objects

Table 5.20: Hierarchy of cluster creepers (Wurmbrand 2005: 275)

In general, Wurmbrand's hierarchy is visible in the data: While SVPs are by far the most common type of cluster creepers, the data also include interruption by adverbial modifiers, idioms (i.e. the phrasal SVPs) and bare nouns. Of the eight NP objects (OBJ1), there are five indefinite and three definite instances. What Wurmbrand's hierarchy does not include, however, are the instances of adposition stranding as illustrated in (5.75). The treebank data suggest that a hierarchy of cluster creepers for Dutch should include this type of cluster creeping as well (between SVPs and adverbs/idioms/bare nouns).

Multiple cluster creepers

The 14 constructions that are not included in Table 5.19 form a heterogenous group that is not encountered in the literature on cluster creeping. Those examples contain multiple cluster creepers. It is hard to draw any generalizations about this kind of constructions. Out of the 14 instances, 10 cluster creepers consist of a modifier, combined with a direct object, a predicative complement, a prepositional complement or a locational/directional complement. With regard to the syntactic category of the complex cluster creepers, any combination of lexical and phrasal categories seems to be possible. Some examples are given in (5.79).

(5.79) a. . . . den dokter heeft eerst moeten *tien minuten die twee vrouwen* kalmeren
. . . the doctor has first must ten minutes those two women calm-down
voor ie het onderzoek kon doen.
before he the examination could do
'. . . The doctor first had to calm down those two women for ten minutes before he could do the examination.' [CGN, fvn400019__191]

b. . . . alhoewel dat ik er wel 'ns zou *graag aan* meedoen.
. . . although that I there indeed some time would gladly on participate
'. . . although I would like to participate in that.' [CGN, fvb400165__191]

c. . . . als je zeg maar homo bent en dan uh ja gewoon nie ja je weet
. . . if you say but gay are and then uh yeah just not yeah you know
niet hoe je het met je ouders moet *'t erover* hebben . . .
not how you it with your parents must it there-over have . . .
'. . . for example if you are gay and you don't know how you should talk about it with your parents.' [CGN, fna000541__298]

In (5.79a) the cluster contains a temporal modifier and a direct object NP. (5.79b) is a combination of an adverbial modifier and adposition stranding. In (5.79c) not only the preposition occurs within the cluster, but the PC as a whole is realised in situ. Moreover, the cluster is interrupted by the direct object as well. Not surprisingly, all instances of such *complex* creeping constructions occur in the spoken data (CGN).

Position of the cluster creepers

Another aspect regarding cluster creeping is the position of the non-verbal elements. In section 1.6.2 it was said that in clusters with more than two verbs, the non-verbal element typically occurs right in front of the main verb.

In the data set, there are 32 cases of cluster creeping in constructions with three or four verb forms. In 20 cases, the cluster creeper occurs just in front of the main verb, as in (5.80a), whereas in 12 constructions, they occupy a more leftward position in the cluster, as in (5.80b). The numbers confirm the statement of Haeseryn et al. (1997), but the amount of relevant examples in the treebanks is very low.

(5.80) a. . . . iemand die zich heeft weten *binnen* te werken in kringen met een
. . . someone who himself has know in to work in circles with a
hoog sociaal aanzien . . .
high social standing . . .
'. . . someone who has managed to work his way up into high society . . . ' [LASSY, dpc-ind-001652-nl-sen.p.11.s.1]

b. dus dat huisje wat we daar hebben *neer* laten zetten . . .
so that house.DIM what we there have down let put . . .
'so that little house that we got built over there.' [CGN, fni007330__43]

Language-internal variation

Haeseryn et al. (1997) state that cluster creeping is more typical in Belgian Dutch compared to Netherlandic Dutch. Since CGN contains meta-information on the origin of the data, it is possible to verify that aspect in the treebank results as well. Out of the 194 occurrences of cluster creeping in CGN, 146 constructions are part of the Belgian data set, while the remaining 48 constructions occur in the Netherlandic data, so the data indeed show that cluster creeping is more common in Belgian Dutch. In section 4.2 it was mentioned that CGN contains twice as much Netherlandic data as Belgian data. If we normalize the data, it turns out that cluster creeping occurs six times more often in the Belgian data compared to the Netherlandic part of the corpus.

A note on separable verbs in LASSY

As mentioned in section 5.5.1, separable verbs may be written as one word if the SVP occurs next to the verb. In those cases the SVPs are not individually tagged in the treebanks. It is possible, however, to detect the clusters containing an SVP by extracting the root forms of the verbs in the clustering constructions in LASSY. In the `root` tag of the verb the root and the SVP are separated by an underscore, e.g. bel_op for the verb *opbellen* 'call'. The numbers are given in Table 5.21.

	MV $\neq 1$	MV $= 1$	SUM
Separable verbs	**2,618**	406	3,024
Non-separable verbs	6,547	2,008	8,555
SUM	9,165	2,414	**11,579**

Table 5.21: Distribution of separable verbs within clusters in LASSY

The 2,618 constructions in which the main verb is not the first verb in the clusters include 2,415 constructions with cluster creepers, as in (5.81a-b),[27] and 146 constructions without cluster creeping in which the SVP occurs in front of the verb cluster. as in (5.81c).

(5.81) a. Dit interview vond plaats voor dat werd bekendgemaakt ...
this interview found place before that was known-made ...
'This interview took place before that was announced ...' [LASSY, dpc-cam-001283-nl-sen.p.11.s.7]

b. Ian Strachan is een Londense clubber die zich modieus weet uit te dossen ...
Ian Strachan is a London clubber who himself fashionably knows out to dress ...
'Ian Strachan is a London clubber who knows how to dress fashionably ...' [LASSY, dpc-ind-001652-nl-sen.p.9.s.1]

c. ... als hij de slavernij definitief af zou schaffen ...
... if he the slavery definitively off would abolish
'... if he would abolish slavery for good ...' [LASSY, WR-P-E-I-0000050394.p.9.s.185]

The results show that in constructions with separable verbs cluster creeping is a frequently occurring phenomenon, in contrast to the observations in Table 5.19. The creeping constructions account for 20.86% of all clustering constructions in LASSY, and for 79.86% of all separable verbs in clusters in LASSY.

Note that separable verbs are only represented as such in the root forms but not in the lemmas. It is possible to retrieve SVPS in this way in the LASSY treebank, but not in CGN, since the CGN treebank only includes lemmas but no root forms. It is thus not possible to compare the results in Table 5.21 to the frequency of separable verbs in CGN. Exploring alternative ways of retrieving those constructions in CGN remains future work.

[27] The results include the examples with the separately tagged SVPS as well.

5.5.3 Conclusion

The case study presented in this section investigated the occurrence of cluster creepers in the CGN and LASSY treebanks. Compared to the large amount of clustering constructions, the treebanks show that cluster creeping is a low-frequent phenomenon in Dutch, except in the case of SVPS. Despite the small set of treebank results, the variety of the creeper types turned out to be rather large. All categories mentioned in Haeseryn et al. (1997) are included in the data. Moreover, a subset of the cluster creepers consists of a combination of several creeper types. Those constructions are not mentioned in the literature on the phenomenon, showing that corpus-based research can add additional insights into linguistic phenomena.

Further work is needed on how to deal with the inconsistent spelling in Dutch regarding separable verb particles, as well as with the problematic annotation of separable verbs in the treebanks. It would also be interesting to investigate the phenomenon in a larger corpus, for example in the SoNaR treebank (Oostdijk et al. 2013).

5.6 Conclusion

This chapter presented the treebank-based investigation of verb clusters and phenomena that are related to verb clusters, such as the IPP effect and cluster creeping. It was shown that the treebank instances of clustering verbs are a subset of the verbs that select a VC complement. While it was relatively straightforward to extract verb clusters without a *te*-infinitive, clusters with *te*-infinitives were not trivial to identify. On the one hand, this can be attributed to the fact that predicative *te*-infinitives and *te*-infinitives that are a part of a fixed expression also received a VC tag in the treebanks. On the other hand, it is not always easy to distinguish constructions in which a *te*-infinitive is selected in the second pole from constructions in which it is selected in the Nachfeld.

Table 5.22 provides an overview of the clustering verbs detected in the treebanks, taking into account the choice of their verbal complement (i.e. past participle, bare infinitive, or *te*-infinitive), and their occurrence in IPP constructions. The figures in the CLUSTERING column are the sums of the constructions in which the verbs unambiguously select a verbal complement in the cluster. These include the sums of the figures in Table 5.7 (22,992 hits), and the figures in Table 5.14 for the verbs selecting a *te*-infinitive in the cluster, i.e. all constructions in which obligatory IPP verbs select

a *te*-infinitive (374 hits), and the constructions in which optionally IPP verbs selecting a *te*-infinitive appear as IPP (16 hits). This amounts to 23,382 verbs selecting their VC in the cluster. The figures for the IPP constructions are the results presented in Table 5.16 (1,101 hits). Those constructions are a subset of the clustering constructions.

		LEMMA	VC TYPE	CLUSTERING	IPP
SUBJECT-ORIENTED VERBS	NO IPP	*worden* 'be'	PSP	5,190	–
		hebben 'have'	PSP	3,002	–
			INF	177	–
		zijn 'be'	PSP	2,968	–
			INF	92	–
	OBLIGATORY IPP	*kunnen* 'can'	INF	3,030	213
		moeten 'have to '	INF	2,232	191
		zullen 'will'	INF	2,035	1
		mogen 'may'	INF	365	29
		(be)hoeven 'have to'	TE INF	58	3
			INF	9	4
		dienen 'have to'	TE INF	34	1
			PSP	1	–
		(be)horen 'ought to'	TE INF	3	0
		lijken 'seem'	TE INF	39	0
		blijken 'turn out'	TE INF	26	0
		schijnen 'appear'	TE INF	6	0
		heten 'be reputed to'	TE INF	2	0
		gaan 'go, will'	INF	1,470	188
		blijven 'remain, continue'	INF	216	29
		komen 'come'	INF	154	20
			TE INF	41	7
			PSP	1	–
		zitten 'sit '	INF	66	44
			TE INF	53	0
			PSP	13	–
		staan 'stand'	TE INF	27	0
			INF	26	15
			PSP	25	–
		wezen 'be in the process of'	INF	15	13

continued on next page

		LEMMA	VC TYPE	CLUSTERING	IPP
SUBJECT-ORIENTED VERBS	OBLIGATORY IPP	*lopen* 'walk'	INF	7	5
			TE INF	5	0
		liggen 'lie'	PSP	7	–
			INF	1	0
			TE INF	10	0
		willen 'want'	INF	946	53
		weten 'know (how to)'	TE INF	47	19
		zien 'manage'	TE INF	20	0
		plegen 'tend'	TE INF	3	0
	OPTIONAL IPP	*beginnen* 'start, begin'	INF	25	8
			TE INF	4	4
		leren 'learn'	INF	53	19
			TE INF	0	0
		durven 'dare'	INF	22	5
			TE INF	3	3
		proberen 'try'	TE INF	8	8
			INF	4	1
		trachten 'try'	TE INF	1	1
OBJECT-ORIENTED VERBS	NO IPP	*hebben* 'have'	INF	51	–
		krijgen 'get'	PSP	38	–
		vinden 'think, find'	INF	5	–
	OBLIGATORY IPP	*laten* 'let'	INF	536	154
		doen 'do'	INF	70	6
		zien 'see'	INF	93	36
			PSP	5	–
		horen 'hear'	INF	28	19
		voelen 'feel'	INF	2	0
	OPTIONAL IPP	*helpen* 'help'	INF	10	1
			TE INF	0	0
		leren 'teach'	INF	2	1
			TE INF	0	0
Sum				**23,382**	**1,101**

Table 5.22: Treebank-based typology of clustering verbs

The verbs *weigeren* 'refuse', *wensen* 'wish', *hopen* 'hope', *beloven* 'promise', *dreigen* 'threaten', *menen* 'mean', *pogen* 'try', and *wagen* 'dare' occur as the selector of a *te*-infinitive in the treebanks, but not as IPP verbs (see Table 5.14). As discussed in section 5.4, data from the Web confirm that those verbs optionally occur as IPP and thus may occur in verb clusters. However, for some verbs only a couple of such constructions could be found, indicating that some speakers might not use those verbs in IPP constructions. A corpus investigation on larger corpora and/or treebanks might shed more light on the IPP status of those verbs. This is left for future research.

Part III

Analysis

CHAPTER 6

A new treatment of verb clusters

A common property of the argument inheritance proposals discussed in chapter 3 is that subjects and complements are raised in the same way: They are both integrated in the SUBCAT list, c.q. the ARG-ST list of the selecting verb.[1] This is in fact the reason why argument inheritance is also known as *generalized raising*.

In this chapter, it will be argued that complement raising should be differentiated from subject raising, at least for Dutch. The problems regarding generalized raising will be pointed out in section 6.1. An alternative analysis which treats complement raising and subject raising as separate phenomena, will be described in section 6.2. It will be demonstrated that the new analysis is applicable to instances of genuine verb clusters, the third construction, as well as constructions that are ambiguous between clustering and non-clustering. Section 6.3 presents the constraints on complement raising.[2] Section 6.4 addresses the phenomenon of *cluster creeping*, i.e. cases in which complement raising is optional. In section 6.5 the word order variation in Dutch verb clusters will be modelled, employing binary-branching structures. Section 6.6 concludes.

6.1 Why differentiate complement raising from subject raising

This section argues why complement raising ought to be differentiated from subject raising in Dutch. The evidence comes from three sources. They concern the occur-

[1] Except in van Noord & Bouma (1997) and Bouma & van Noord (1998), cf. infra.

[2] Previous versions of sections 6.1 to 6.3 appeared in Van Eynde & Augustinus (2013, 2014).

rence of complement raising with subject control verbs, the binding properties of object raising verbs, and the interaction of complement raising and the passive lexical rule.

6.1.1 Interaction with subject control verbs

In section 3.3.2 it was already pointed out that the argument inheritance principle in (3.47), repeated here in (6.1), allows for the occurrence of subject raising without complement raising, since $\boxed{A}$ may be the empty list.

(6.1)
$$\left[\begin{array}{l}\textsc{arg-st}\ \left\langle \boxed{1} \right\rangle \oplus \boxed{A} \oplus \left\langle \left[\begin{array}{l}\textsc{local}\,|\,\textsc{cat}\ \left[\begin{array}{ll}\textsc{head} & \textit{verb}\\ \textsc{subj} & \left\langle \boxed{1} \right\rangle\\ \textsc{comps} & \boxed{A}\end{array}\right]\end{array}\right] \right\rangle\end{array}\right]$$

What (6.1) does not allow, however, is the occurrence of complement raising without subject raising: The SUBJ list of the selected verb is required to contain one *synsem*, and that *synsem* must be identical to the first argument of the selecting verb.

This constraint is too strict, since there is empirical evidence that complement raising also occurs with subject control verbs, such as *willen* 'want' and *proberen* 'try' in (6.2). The control verbs are in bold and the raised complements in italics.

(6.2) a. Kasparov beschuldigde Gorbatsjov ervan dat hij *het bloedvergieten* niet had
Kasparov accused Gorbachev there-of that he the bloodshed not had
willen stoppen.
want.IPP stop.INF
'Kasparov accused Gorbachev that he had not wanted to stop the bloodshed.' [LASSY, dpc-ind-001648-nl-sen.p.19.s.6]

b. ... nadat ze *zowel de PS als de PRL te vriend* had **proberen** te houden.
... after she both the PS and the PRL as friend had try.IPP to keep.INF
'... after she had tried to keep both the PS and the PRL as an ally.' [LASSY, WR-P-P-I-0000000106.p.7.s.6]

Notice that the control verbs in these sentences are affected by the IPP phenomenon. Several of those verbs also allow complement raising if they do not appear as an IPP verb, as illustrated in (6.3), which is an instance of the third construction (den Besten et al. 1988; den Besten & Rutten 1989).

(6.3) ja en en ik heb *'r* **geprobeerd** te bellen maar d'r werd niet
yes and and I have her try.PSP to call.INF but there was not
opgenomen...
picked-up...
'yes and and I've tried to call her but there was no reply' [CGN, fna000583 __351]

Summing up, subject control verbs are obviously not subject raisers, but they do allow complement raising, both in IPP constructions and in the third construction.

Since the argument inheritance constraint in (6.1) does not subsume the subject control verbs, a separate constraint is needed to model complement raising in clauses like (6.2–6.3):

(6.4)
$$\left[\text{ARG-ST}\ \left\langle \text{NP}\boxed{1}\right\rangle \oplus \boxed{A} \oplus \left\langle \left[\text{LOCAL}\,|\,\text{CAT}\ \left[\begin{array}{l}\text{HEAD}\ \ \textit{verb}\\ \text{SUBJ}\ \ \left\langle \text{NP}\boxed{1}\right\rangle\\ \text{COMPS}\ \ \boxed{A}\end{array}\right]\right]\right\rangle\right]$$

Complying with the way in which subject control verbs are canonically differentiated from subject raising verbs (Pollard & Sag 1994; Sag et al. 2003), as explained in section 3.2, this constraint requires the unexpressed subject of the selected verb to share its index with the first argument of the selecting verb, rather than its entire *synsem* value. This addition of an extra constraint is not a problem as such, but it does raise the suspicion that we are missing a generalization.

6.1.2 Interaction with the binding principles

In HPSG, the binding principles are canonically defined in terms of obliqueness relations in the ARG-ST list (Pollard & Sag 1994; Sag et al. 2003).

> Principle A: An anaphoric pronoun must be coindexed with a less oblique argument on the same ARG-ST list.
>
> Principle B: A non-anaphoric NP may not be coindexed with a less oblique argument on the same ARG-ST list.

Assuming that raised subjects are integrated in the ARG-ST list of the selecting verb, this makes the right prediction for the object raiser *ziet* 'sees' in (6.5).

(6.5) a. dat hij$_i$ zich$_{i/*j}$ die wedstrijd niet meteen **ziet** winnen.
that he$_i$ himself$_{i/*j}$ that game not immediately sees win
'that he does not expect himself to win that game rightaway.'

b. dat hij$_i$ hem$_{j/*i}$ die wedstrijd niet meteen **ziet** winnen.
that he$_i$ him$_{j/*i}$ that game not immediately sees win
' that he doesn't expect him to win that game rightaway.'

The raised reflexive pronoun *zich* 'himself' in (6.5a) must be coindexed with the subject of *ziet* 'sees', yielding the interpretation that he does not expect himself to win the contest. Similarly, the raised personal pronoun *hem* 'him' in (6.5b) cannot be coindexed with the subject of *ziet*, yielding the interpretation that he does not expect that person to win the contest.

Raised subjects thus behave as *bona fide* arguments of the matrix verb, as illustrated by the ARG-ST list of *ziet* for the sentences in (6.5):

(6.6) a. *ziet*: ARG-ST <NP$_i$, 1 NP$_{i/*j}$, (...,) V[SUBJ <1>]>

b. *ziet*: ARG-ST <NP$_i$, 1 NP$_{j/*i}$, (...,) V[SUBJ <1>]>

Raised complements, by contrast, show the opposite behaviour.

(6.7) a. * ... dat hij$_i$ ons *zich*$_i$ niet meteen **ziet** uitschakelen.
... that he$_i$ us himself$_i$ not immediately sees eliminate

b. ... dat hij$_i$ ons *hem*$_{i/j}$ niet meteen **ziet** uitschakelen.
... that he$_i$ us him$_{i/j}$ not immediately sees eliminate
'... that he doesn't expect us to eliminate him rightaway.'

If the raised reflexive pronoun in (6.7a) is integrated in the ARG-ST list of *ziet* and coindexed with its subject, as in (6.8a), then it complies with binding principle A, but the sentence is nonetheless ill-formed. Conversely, if the raised personal pronoun in (6.7b) is integrated in the ARG-ST list of *ziet* and coindexed with its subject, the interpretation is impeccable but in that case it violates binding principle B, cf. (6.8b).

(6.8) a. *ziet*: ARG-ST <NP$_i$, 1 NP , 2 NP$_i$, V[SUBJ <1> , COMPS <2>]>

b. *ziet*: ARG-ST <NP$_i$, 1 NP , 2 NP$_{i/j}$, V[SUBJ <1> , COMPS <2>]>

As a consequence, either the binding principles need to be adapted, or the raised complements should be treated in another way than the raised subjects, integrating the latter but not the former in the ARG-ST list of the selecting verb.

Something along this lines is also pointed out by Pollard (1994: 285–286) and van Noord & Bouma (1997: 231–233). While Pollard (1994) observes that also in

German the uniform attraction analysis of the complements of object raisers poses a problem if one wants to maintain the binding principles, he leaves the issue unsolved. Van Noord & Bouma (1997) solve the issue by stating that the inherited complements appear on the COMPS list of the selecting verb, but not on its ARG-ST list.[3] While this indeed solves the problem with respect to the binding principles, it violates the argument realization principle, which defines the ARG-ST list of a word as the append of the SUBJ and COMPS list, cf. (3.22), repeated in (6.9).

(6.9) Argument Realization Principle (ARP):

$$\textit{word} \Rightarrow \begin{bmatrix} \text{SYNSEM}\,|\,\text{LOC}\,|\,\text{CAT} \begin{bmatrix} \text{SUBJ} & \boxed{A} \\ \text{COMPS} & \boxed{B} \end{bmatrix} \\ \text{ARG-ST}\ \boxed{A} \oplus \boxed{B} \end{bmatrix}$$

6.1.3 Interaction with the passive lexical rule

HPSG canonically treats passivization in terms of a lexical rule which reshuffles the order of the arguments on the ARG-ST list. For Dutch, it can be formalized as the rule in (6.10), after Sag et al. (2003: 313).

(6.10)
$$\begin{bmatrix} \textit{tv-lxm} \\ \text{PHON}\ \boxed{A} \\ \text{ARG-ST}\ \left\langle \text{NP}_i \right\rangle \oplus \boxed{B} \end{bmatrix} \Longrightarrow_{LR} \begin{bmatrix} \text{PHON}\ \ \text{F}_{psp}\left(\boxed{A}\right) \\ \text{SYNSEM}\,|\,\text{LOC}\,|\,\text{CAT}\,|\,\text{HEAD}\,|\,\text{VFORM}\ \ \textit{passive} \\ \text{ARG-ST}\ \ \boxed{B} \ \oplus\ \left\langle \left(\text{PP}_i\right) \right\rangle \end{bmatrix}$$

This rule relates a transitive verbal lexeme to its participial form, fixing the VFORM value to *passive* and changing the order in the ARG-ST list: The second argument of the verbal lexeme becomes the first argument of its passive counterpart.

Assuming that raised subjects are integrated in the ARG-ST list of the selecting verb, this makes the right prediction for the object raising verb *expect* in (6.11).

(6.11) a. We expect them to leave tomorrow.

b. They are expected to leave tomorrow.

Since the noun phrase which is realized by *them* is the second argument of the lexeme *expect*, it can become the first argument of its passive counterpart *expected*.

Raised complements, by contrast, behave differently, as illustrated in (6.12).

[3] For an example of this, see the lexical entry of *laten* 'let' in (3.68).

(6.12) a. . . . dat hij *het huis* **probeerde** te verkopen.
. . . that he the house tried to sell
'. . . that he tried to sell the house.'

b. * . . . dat *het huis* **werd geprobeerd** te verkopen.
. . . that the house was tried to sell

The italicized complement of *verkopen* 'sell' in (6.12a) is raised and realized as a dependent of the subject control verb *proberen* 'try', but in contrast to the raised subject in (6.11) it cannot become the first argument of the passive *geprobeerd* 'tried'. As a consequence, if passivization is treated in terms of reshuffling along the lines of (6.10), then the raised complements should not be integrated in the ARG-ST list of the selecting verb.

Note that in German constructions equivalent to (6.12a) are considered grammatical, at least for a number of speakers. Such constructions are known as *long passive* or *remote passive*, see amongst others Kiss (1995: 136–140) and Müller (2002: 136–138). An example is given in (6.13), cited from Müller (2002: 136).

(6.13) weil der Wagen oft zu reparieren versucht wurde.
because the car.NOM often to repair tried was
'because many attempts were made to repair the car.'

The object of the embedded verb *reparieren* 'repair' is realized as the subject of *versuchen* 'try', which is indicated by the fact that it gets nominative case.

While the passivization of raised complements is be allowed (to some extent) in German, such constructions are ungrammatical in Dutch.

6.2 An alternative treatment of complement raising

This section presents an alternative analysis for argument inheritance, which differentiates complement raising from subject raising. It will be shown how the problems discussed in section 6.1 can be avoided by separating those two mechanisms,.

6.2.1 Complement raising versus subject raising

In order to avoid the problems with the binding principles and the passive lexical rule mentioned in section 6.1, raised subjects should be integrated in the ARG-ST list of the selecting verb, while raised complements should not. In that way, the canonical analyses of binding and the passive can be kept as they are.

For the treatment of subject raising this implies that the same lexical constraints can be used as those for English, i.e. one for subject raisers and one for object raising verbs, see (3.24) and (3.26) in section 3.2.4, repeated as (6.14) and (6.15) (Ginzburg & Sag 2000: 22).

(6.14) *s-rsg-lx* ⇒ $\left[\text{ARG-ST} \left\langle [\text{LOC } \boxed{1}] , [\text{SUBJ } \langle [\text{LOC } \boxed{1}] \rangle] \right\rangle\right]$

(6.15) *o-rsg-lx* ⇒ $\left[\text{ARG-ST} \left\langle \text{NP} , [\text{LOC } \boxed{1}] , [\text{SUBJ } \langle [\text{LOC } \boxed{1}] \rangle] \right\rangle\right]$

For the treatment of complement raising, by contrast, the phrasal constraint in (6.16) can be used.

(6.16) *hd-ph* ⇒
$$\begin{bmatrix} \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{COMPS} & \boxed{A} \bigcirc \boxed{B} \\ \text{HEAD-DTR} \mid \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{COMPS} & \boxed{A} \\ \text{NONHD-DTR} \mid \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{COMPS} & \boxed{B} \end{bmatrix}$$

The *Complement Raising Principle* (CRP) states that in a headed phrase, the COMPS list of the non-head daughter is added to the COMPS list of the mother.[4] $\boxed{B}$ may be empty, but it may also contain one or more members.

Cancellation of elements from the COMPS list is modelled in the definition of the phrases of type *head-complement*, which was already defined in (3.12) and repeated here as (6.17).

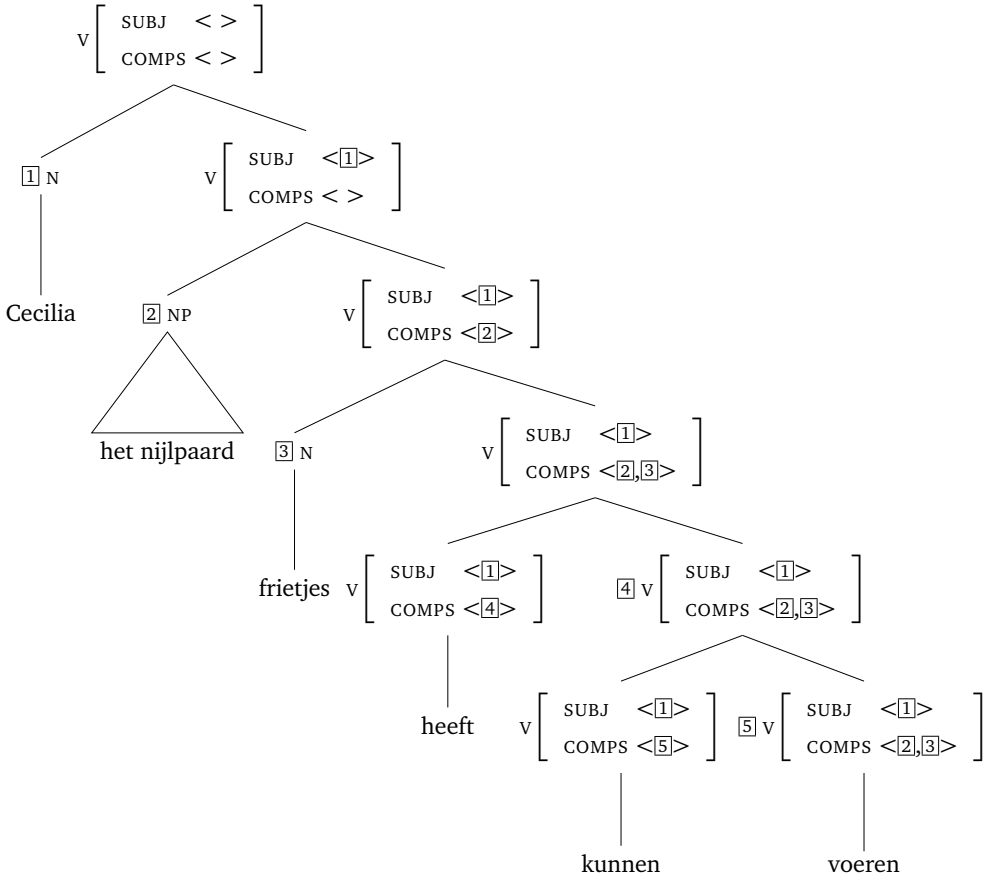
(6.17) *hd-comp-ph* ⇒
$$\begin{bmatrix} \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{COMPS} & \boxed{A} \\ \text{HD-DTR} \mid \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{COMPS} & \boxed{A} \oplus \left\langle \boxed{1} \right\rangle \\ \text{NONHD-DTR} \mid \text{SYNSEM} & \boxed{1}\ \textit{synsem} \end{bmatrix}$$

Since *head-complement-phrase* is a subtype of *headed-phrase*, it follows that the COMPS list can shrink and expand at the same time.

Applying (6.16) and (6.17) to (3.45) yields the tree structure in (6.18).

[4] In non-headed phrases, such as coordinate structures, the COMPS list of the mother is identical to the COMPS lists of each of the conjunct daughters separately. In *he buys and sells cars*, for instance, the coordinate phrase *buys and sells* has the same COMPS list as its conjunct daughters, *buys* and *sells*.

(6.18)



The modal *kunnen* 'can' shares its subject requirement with the infinitival complement (1), but not its COMPS list (<2,3>). The latter is propagated directly from the non-head daughter to the mother as stated in (6.16). The same holds for the perfect *heeft* 'has': It inherits the SUBJ list of its infinitival complement, but not its COMPS list. The infinitival complements of *kunnen* (5) and *heeft* (4) are cancelled off as defined in (6.17). The constraint on head-complement phrases furthermore deals with the cancellation of the non-verbal complements: The non-verbal complements are cancelled off the COMPS list in a stepwise way, from the most (*frietjes* 'chips') to the least oblique (*het nijlpaard* 'the hippo'), giving rise to a construction with cross-serial dependencies. Finally, the combination of the saturated VP with the subject is modelled by means of the constraint on head-subject phrases that was presented in (3.15) in section 3.2.2.

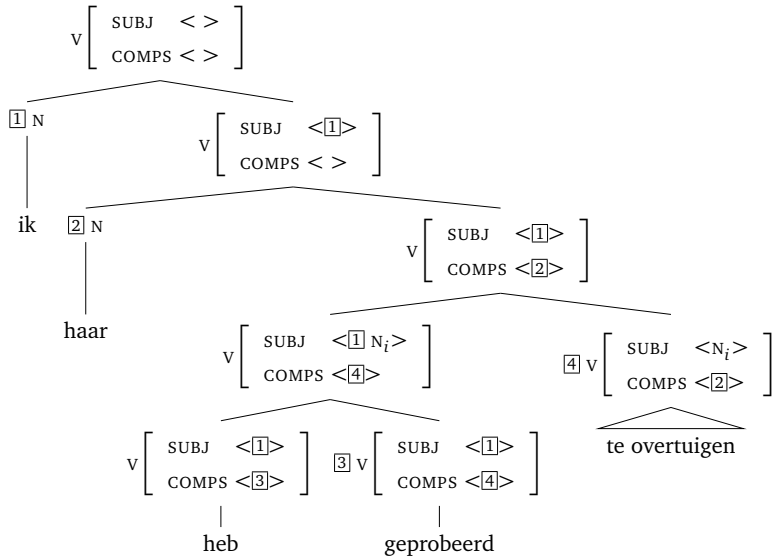
Notice that the analysis of complement raising presented here shows some similarity to the CG and GPSG analyses presented in section 3.1. In the CG analysis as shown in (3.2) the complement *de nijlpaarden* 'the hippos' is not inherited by the se-

lecting verbs *helpen* 'help' and *zien* 'see', but directly propagated to the dominating node. A similar mechanism can be observed in Johnson's GPSG analyses. In the tree representation in (3.5b), the features NP_{dat} and NP_{acc} are not inherited by *können* 'can', but directly propagated to the mother node.

The adaptation of a uniform argument inheritance principle to a treatment which differentiates complement raising from subject raising provides exactly what is needed to avoid the problems mentioned in the previous section: It allows for complement raising in cases where there is no subject raising, and it does not integrate the raised complements in the ARG-ST list of the selecting verb. (6.20) illustrates the analysis of complement raising out of the Nachfeld in (6.19), i.e. an instance of the third construction.

(6.19) ... dat ik haar heb geprobeerd te overtuigen.
... that I her have tried to persuade
'... that I have tried to persuade her.'

(6.20)



The fact that the past participle *geprobeerd* 'tried' selects its complement in the Nachfeld and not in the cluster is modelled by means of a different branching structure (see also section 6.4 and 6.5 for a discussion of such constructions). This poses no problems for the treatment of complement raising: First *heb* 'have' and *geprobeerd* form a cluster. The unsaturated COMPS list of *geprobeerd* is propagated up the tree, and cancelled off after the combination with *te overtuigen* 'to persuade'. The unsaturated COMPS list of *te overtuigen* is propagated up the tree, while its subject shares its

index with the subject of *heb geprobeerd* according to the lexical constraint on subject control verbs in (3.32).

If *geprobeerd* 'tried' in (6.19) would appear as an infinitive (*proberen* 'try'), i.e. as an IPP verb, the analysis will be similar to (6.20), i.e. another case of complement raising without subject raising, but the branching structure would be like the construction in (6.18), since *proberen* is a clustering verb if it appears as IPP.

6.2.2 Complement raising versus complement extraction

Complement raising need not only be differentiated from subject raising, but also from complement extraction. The latter concerns a long distance dependency that may cross clause boundaries, as in (6.21–6.22).

(6.21) Who [do you think [she said [she would date]]]?

(6.22) Wie [beweert ze [dat ze in Parijs ontmoet hebben]]?
who claims she that they in Paris met have
'Who does she claim they met in Paris?'

The complements of *date* and *ontmoet* 'met' are extracted and realized as a filler of the main clause. In HPSG, this is modelled in terms of a lexical rule which subtracts elements from the COMPS list and adds them to the non-local SLASH list, see (Ginzburg & Sag 2000).

Complement raising, by contrast, is a medium-distance dependency which does not cross clause boundaries. More specifically, complements cannot be raised beyond the first pole.[5] That complementizers are a boundary for complement raising is illustrated in (6.23).

(6.23) a. ...dat ze beweert **dat** ze *hem* in Parijs ontmoet hebben.
...that she claims that they him in Paris met have
'...that she claims that they met him in Paris.'

b. * ...dat ze *hem* beweert **dat** ze in Parijs ontmoet hebben.
...that she him claims that they in Paris met have

The italicized complement of *ontmoet* 'met' cannot be raised out of the clause that is introduced by the complementizer *dat* 'that'.

That finite verbs are a boundary for complement raising is less obvious, since it is possible to realize the complement of the main verb in the Vorfeld, as in (6.24).

[5]The first pole is the position that is taken by the complementizer in verb-final clauses and by the finite verb in verb-initial clauses, i.e. verb-first and verb-second clauses, see section 1.1.

(6.24) Dat boek zal Peter toch niet kunnen vinden.
that book shall Peter anyway not can find
'That book, Peter will not be able to find it anyway.'

Notice, though, that this is an instance of topicalization, and that topicalization is canonically treated as a long-distance dependency in HPSG, amongst others because it can cross clause boundaries, as in (6.25).

(6.25) That man I wish I had never known.

A useful test for differentiating topicalization from complement raising in Dutch is exemplified by the contrast in (6.26).

(6.26) a. Hij zal jou/je toch niet kunnen vinden.
He shall you/you.RED anyway not can find
'He will not be able to find you anyway.'

b. Jou/*je zal hij toch niet kunnen vinden.
you/*you.RED shall he anyway not can find
'You he will not be able to find anyway.'

Pronominal complements can be raised out of a verb cluster, as in (6.26a), no matter whether they take the full form or a phonologically reduced form, i.e. a form with a mute vowel or without a vowel. Extraction, by contrast, as in (6.26b), is only possible for the full form (Van Eynde 1999).[6]

The medium-distance nature of complement raising is thus clear from the fact that it cannot go beyond the first pole: It is bounded by the complementizer in verb-final clauses and by the finite verb in verb-initial clauses. How these constraints can be spelled out in formal terms is discussed in section 6.3.

6.3 Constraints on complement raising

Section 6.2 showed that the CRP is a rather powerful device. In some Dutch constructions complement raising is not allowed, however. In order to prevent overgeneration a number of constraints on complement raising need to be added. Those constraints can be formulated in a way similar to that in which complement raising is (entirely) blocked in English.

[6]This restriction holds for extracted *complements*. Subjects may always occur in the Vorfeld, no matter whether they are full forms or reduced forms.

6.3.1 How to block complement raising

English is a language that allows subject raising and complement extraction, but it does not allow complement raising.[7] Within HPSG, this is formalized as the Empty COMPS Constraint (ECC), as defined in Ginzburg & Sag (2000: 33).

(6.27) Empty COMPS Constraint

$$\textit{phrase} \quad \Rightarrow \quad \left[\text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{COMPS} \quad \left\langle \right\rangle \right]$$

If phrases are required to have an empty COMPS list, then it follows that complement raising is blocked.

The fact that English allows adposition stranding does not provide any evidence against this assumption, since the stranding typically results from complement extraction, as in (6.28).[8]

(6.28) a. *What* do you think they were talking **about**?

b. *This* I would never dare talk **about** in her presence.

Stranding that results from complement raising, as in (6.29), is not possible.

(6.29) a. * John heard *this* us talk **about**.

b. * We saw him *that* give a talk **about**.

In that respect, English differs from Dutch, where the equivalents of (6.29) are well-formed.

(6.30) a. Jan heeft *er* ons **over** horen praten.
Jan has there us about hear talk
'Jan heard us talk about it.'

b. We hebben hem *daar* een lezing **over** zien geven.
we have him there a talk about see give
'We saw him give a talk about that.'

[7] This can be considered as yet another argument to differentiate subject raising from complement raising.

[8] English also shows adposition stranding in pseudo-passives, e.g. *This bed was slept in.* Similar to the analysis of canonical passive constructions in HPSG, such constructions can be dealt with by means of a lexical rule, see for instance Kim (2009). This type of adposition stranding does not occur in Dutch.

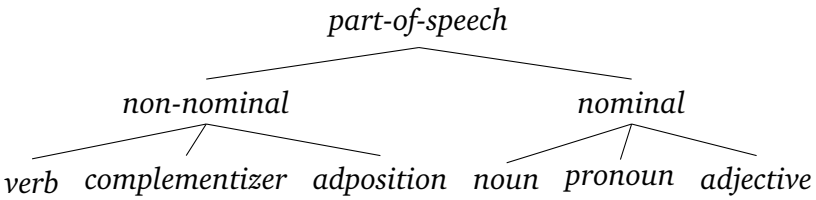
In sum, the addition of the Empty COMPS Constraint suffices to rule out complement raising from a language.

Dutch obviously does not abide by the ECC, but this does not mean that its complements can be raised anywhere. In order to prevent complement raising in cases where it is not allowed, constraints on the CRP should be added stating that the COMPS list should be empty in those constructions. The constraints on complement raising in Dutch will be spelled out in the remainder of this section.

6.3.2 No complement raising beyond the first pole

In section 6.2.2 it was shown that complements cannot be raised beyond the first pole. An extra POSITION feature is needed to model the relevant constraint. As will be motivated in the remainder of this section and in chapter 7, it need not only be added to the HEAD values of verbs, but also to the HEAD values of complementizers and adpositions, cf. the hierarchy in (6.31). As in Chomsky (1970) and Jackendoff (1977) verbs and adpositions are considered as non-nominal, i.e. [– N].[9] In addition, the complementizers can be considered as *non-nominal* as well.

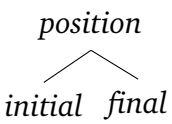
(6.31)



The type declaration of the POSITION feature is presented in (6.32) and its possible values are given in (6.33).

(6.32) *non-nominal*: $\left[\text{POSITION} \quad \textit{position}\right]$

(6.33)



In terms of this dichotomy, the non-finite verbs are invariably *final* and the imperative forms *initial*. The other finite forms can occur in either position, and hence receive the underspecified *position* value.[10] The POSITION feature is added to the HEAD values of

[9] In Chomsky (1970) the lexical categories are analyzed in terms of the Boolean features N and V: Verbs are [–N, +V], nouns are [+N, –V], adjectives are [+N, +V] and adpositions are [–N, –V].

[10] The term *initial* subsumes both verb-first and verb-second.

verbs, so the value of POSITION is shared between the mother and the head daughter. In this way, it is assured that, for instance, clauses are verb-initial if they contain a head verb with the feature [POSITION *initial*].

final	non-finite forms
initial	imperative forms
position	non-imperative finite forms

Table 6.1: Three types of verb forms

As Dutch complementizers are all head-initial, they are all specified as [POSITION *initial*]. This addition makes it possible to use the following constraint in order to block complement raising out of verb-initial VPs and CPs:[11]

(6.34) $$\left[\begin{array}{l} \textit{hd-ph} \\ \text{SYNSEM}\,|\,\text{LOC}\,|\,\text{CAT}\,|\,\text{HEAD}\,|\,\text{POSITION}\ \textit{initial} \end{array}\right] \Rightarrow \left[\text{SYNSEM}\,|\,\text{LOC}\,|\,\text{CAT}\,|\,\text{COMPS}\ \langle\,\rangle\right]$$

Phrases which have the feature [POSITION *initial*] as a HEAD value are required to have an empty COMPS list. This suffices to block complement raising out V-initial VPs and CPs.[12] The constraint in (6.34) models the fact that complements cannot be raised beyond the first pole. As such, it captures what differentiates complement raising from complement extraction.

6.4 Optional versus obligatory complement raising

In the previous sections it was illustrated how the CRP succeeds in explaining the relation between the verbs and their arguments in clustering constructions as well as in the third constuction. For varieties of Dutch that do not allow any non-verbal elements in the cluster, complement raising is obligatory for those phrases headed by an obligatory clustering verb, or by an optional clustering verb appearing as IPP (cf. Table 5.22).

[11] In this analysis complementizers are treated as heads, following Ginzburg & Sag (2000: 46–49). It contrasts with Pollard & Sag (1994: 44–46), in which complementizers are treated as markers. If one adopts the marker analysis of complementizers, one should formulate separate constraints for verb-initial VPs and CPs. In that case, the constraint in (6.34) has to be reformulated, for instance, by requiring that the clausal sister of the complementizer must have an empty COMPS list.

[12] It does not block complement extraction, though, since it does not require the SLASH value to be empty.

The examples from West Flemish discussed in the literature study and the treebank instances with cluster creepers (see section 5.5) indicate that for some varieties of Dutch, complement raising is optional. Some examples with cluster creepers taken from section 5.5 are repeated in (6.35).

(6.35) a. Ik heb mijn agenda niet hoeven *om* te gooien ...
I have my agenda not need over to throw ...
'I did not have to completely change my schedule ...' [LASSY, dpc-rou-000479-nl-sen.p.10.s.14]

b. de dokters zeggen wel dat 't gaat *goed* komen.
the doctors say that it goes good come
'The doctors say that it will be fine.' [CGN, fva400370__6]

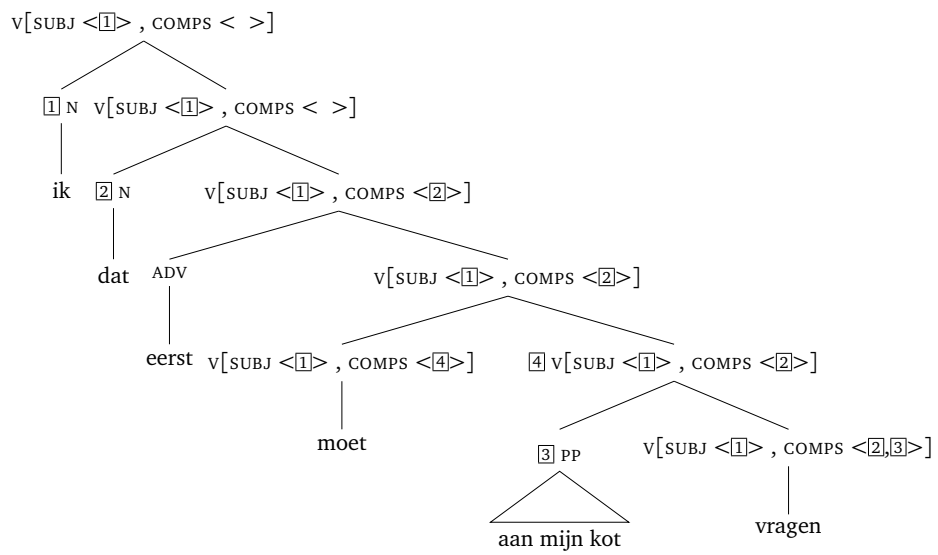
c. ...of dat 'k ik dat eerst moet *aan mijn kot* vragen ...
...or that I I that first must to my student's apartment ask ...
'... or that I should ask (the people of) my student's apartment first.' [CGN, fva400507__4]

d. ...hoe je het met je ouders moet *'t erover* hebben ...
...how you it with your parents must it there-over have ...
'...how you should talk about it with your parents ...' [CGN, fna000541__298]

As was shown in the treebank study, cluster creeping by a separable verb particle, a stranded adposition, a predicative adjective, or a bare noun is not impossible, e.g. (6.35a–b). Moreover, while instances of cluster creeping by a phrase and constructions containing multiple cluster creepers are canonically excluded in Dutch, some treebank examples were found as well (6.35c–d).

The analysis of the constructions in (6.35) is unproblematic for the treatment of complement raising presented in section 6.2. The tree structure for (6.35c) is given in (6.36).

(6.36)

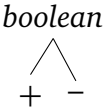


Instead of raising the PP *aan mijn kot* 'to my student's appartment', it is realized in situ and thus cancelled off the COMPS list of *vragen* 'ask' instead of sharing it with the mother node. This yields a partially saturated VP (*moet aan mijn kot vragen*), to which the CRP can be applied to deal with the raising of the nominal *dat* 'that'.

For varieties of Dutch that do not allow cluster creeping, or only allow cluster creeping for a limited set of non-verbal elements, the treatment presented here is too permissive. Cluster creeping can be modelled by defining which elements can appear in the cluster. In order to do this, the Boolean feature CLUSTER is employed. Its type declaration is presented in (6.37) and its possible values are given in (6.38).[13]

(6.37) *category*: $\left[\text{CLUSTER} \quad \textit{boolean}\right]$

(6.38)



The feature is added to the CAT values of objects of type *sign*. The types that obligatorily appear in the cluster, are specified as [CLUSTER +]. This is the case for finite verbs

[13] In chapter 3 it was shown that the modelling of cluster creeping was done by means of the feature NPCOMP in Hinrichs & Nakazawa (1994), by means of LEX in Kathol (2000) and Müller (2002) and by means of the ZONE feature in Bouma & van Noord (1998). While the purpose of the CLUSTER feature is similar, the mechanics for structure sharing are different, which is why another feature name was chosen.

in verb-final postion, bare infinitives, past participles, and cluster creepers that are obligatorily selected in the cluster (i.e. SVPs in some varieties of Dutch). Signs that are specified as [CLUSTER –] can never appear in the cluster. This is the case for finite verbs in verb-initial position, complementizer phrases (e.g. phrases with an *om te* infinitive) and most types of non-verbal phrases, as they never appear in the cluster. A third set of signs has a CLUSTER feature with the underspecified value *boolean*. Those signs include the *te*-infinitives and the cluster creepers that are allowed in and outside the cluster. For Standard and colloquial Dutch those cluster creepers include P-final adpositions, non-subject bare nouns, predicative adjectives and predicative participles.

In all headed phrases, if the CLUSTER feature of the mother has a positive value then the head and non-head daughters are specified as [CLUSTER +], as stated in the constraint in (6.39).

(6.39)
$$\begin{bmatrix} \textit{hd-phrase} \\ \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{CLUSTER} + \end{bmatrix} \Rightarrow \begin{bmatrix} \text{HD-DTR} \mid \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{CLUSTER} + \\ \text{NONHD-DTR} \mid \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{CLUSTER} + \end{bmatrix}$$

If one of the daughters has a negative value, the entire phrase is [CLUSTER –] as well. This is captured by the constraints in (6.40).

(6.40) a.
$$\begin{bmatrix} \textit{hd-phrase} \\ \text{HD-DTR} \mid \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{CLUSTER} - \end{bmatrix} \Rightarrow \begin{bmatrix} \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{CLUSTER} - \end{bmatrix}$$

b.
$$\begin{bmatrix} \textit{hd-phrase} \\ \text{NONHD-DTR} \mid \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{CLUSTER} - \end{bmatrix} \Rightarrow \begin{bmatrix} \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{CLUSTER} - \end{bmatrix}$$

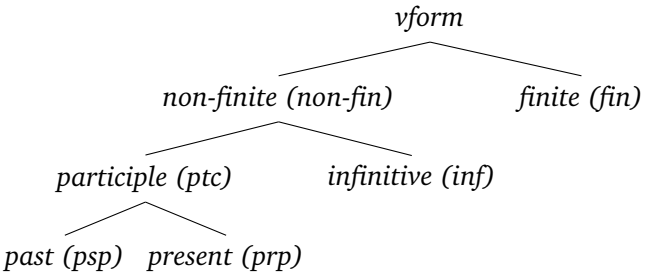
An underspecified value does not influence the cluster value of the mother node. If both the head and non-head daughters are specified as [CLUSTER *boolean*], the entire phrase also receives an underspecified value.

Obligatory clustering verbs share their cluster value with the cluster value of their verbal complement, i.e. they require it to be [CLUSTER +]. This can be modelled by adding the following constraint to the lexical entries of those verbs:

(6.41)
$$\left[\begin{array}{l}\text{SS}\,|\,\text{LOC}\,|\,\text{CAT}\left[\begin{array}{ll}\text{HEAD} & \left[\begin{array}{ll}\textit{verb} & \\ \text{VFORM} & \neg\textit{ptc} \\ \text{POSITION} & \textit{final}\end{array}\right] \\ \text{CLUSTER} & \boxed{1}+ \\ \text{COMPS} & \left\langle \text{V}\left[\text{SS}\,|\,\text{LOC}\,|\,\text{CAT}\,|\,\text{CLUSTER}\ \boxed{1}\right]\right\rangle\end{array}\right]\end{array}\right]$$

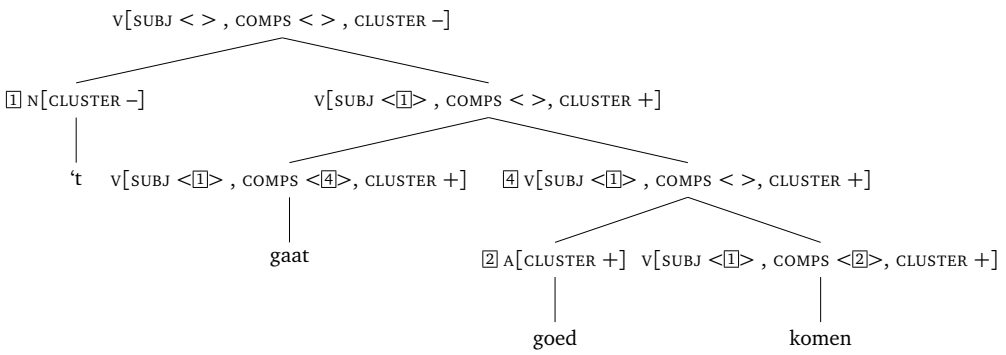
The constraint is applicable to the obligatory clustering verbs (cf. Table 5.22) that appear in verb-final position as a finite verb or a bare infinitive (i.e. not as a participle). The inventory of V(ERB)FORM values is given in (6.42).

(6.42)



Verbs that do not cluster require their verbal complement to be [CLUSTER –], while optionally clustering verbs do not put any requirements on the CLUSTER value of their complement, as they can select a complement inside and outside the cluster. Notice that the cluster value of both clustering and non-clustering verbs is [CLUSTER +] if they appear in verb-final position. Past participles, for instance, have the specification [CLUSTER +], but they require their complement to be [CLUSTER –] if it is a verbal complement, cf. *geprobeerd* 'tried' in (6.20), or [CLUSTER *boolean*] if the complement is a predicative participle, e.g. *vermoord gevonden* 'found murdered' (see also the examples in section 1.10.2). In (6.43) it is shown how the addition of the CLUSTER feature correctly accounts for the cluster creeping of *goed* 'good' in (6.35b).

(6.43)



In (6.43) the verbs are specified as [CLUSTER +], as the construction contains a finite verb and a bare infinitive in verb-final position. The predicative adjective *goed* has an underspecified CLUSTER value, as it may appear inside or outside the verb cluster. Given (6.39), *gaan* 'go' requires its complement to be [CLUSTER +]. Hence, the CLUSTER value of *goed* is contextually resolved as [CLUSTER +]. As subjects are always [CLUSTER –], the entire phrase becomes [CLUSTER –] after the verb cluster is combined with the subject *'t* 'it'. In order to avoid cluster creeping by subjects, the constraint on head-subject phrases can be refined as follows:

(6.44)

$$\textit{hd-subj-ph} \quad \Rightarrow \quad \left[\begin{array}{ll} \textsc{synsem} & \left[\textsc{loc}\,|\,\textsc{cat}\left[\begin{array}{l}\textsc{subj}\ \langle\ \rangle\\ \textsc{cluster}\ -\end{array}\right]\right] \\ \textsc{hd-dtr} & \left[\textsc{synsem}\,|\,\textsc{loc}\,|\,\textsc{cat}\,|\,\textsc{subj}\ \langle \boxed{1} \rangle\right] \\ \textsc{non-hd-dtr} & \left[\textsc{synsem}\ \boxed{1}\left[\begin{array}{l}\textit{synsem}\\ \textsc{loc}\,|\,\textsc{cat}\,|\,\textsc{cluster}\ -\end{array}\right]\right] \end{array}\right]$$

Following from the constraint in (6.40b), a head-subject phrase has a negative CLUSTER value, since its subject daughter is specified as [CLUSTER –].[14]

By specifying which elements can and cannot appear within the verb cluster, one can account for regional, dialectal and even inter-speaker variation of cluster creeping. For instance, dialects that allow cluster creeping by phrasal elements, such as entire PPs, will assign [CLUSTER *boolean*] to the PP *aan mijn kot* in (6.36), and thus allow selection by the obligatory cluster verb *moeten* 'must'. Varieties that do not allow this type of cluster creeping, assign a negative CLUSTER value to PPs, which prevents their occurrence in the cluster.

The introduction of the CLUSTER feature furthermore correctly accounts for the fact that in instances of the third construction the *te*-infinitive is in the Nachfeld if it is selected by a participle (as participles require their verbal complement to be [CLUSTER –]).

Moreover, it provides a uniform analysis for constructions that are in between clustering and non-clustering, i.e. constructions in which an optional IPP verb does not appear in the IPP form, cf. the examples in (5.30) and (5.45), repeated in (6.45).

[14] In order to avoid cluster creeping by an extracted element, a similar constraint can be added to the filler daughter in head-filler constructions.

(6.45) a. ja dan uh als je de grote lijnen 'ns probeert te schetsen...
yeah then uh if you the big lines once try to sketch
'yeah and then if you try to sketch the broad outlines ...' [CGN, fnf007116__4]

b. oh ze zal proberen wakker te worden zegt ze.
oh she will try awake to become says she
'oh she says she will try to be awake.' [CGN, fnc008001__218]

c. en dan moet je dat proberen alleen maar te doen in win-winsituaties.
and then must you that try only just to do in win-win situations
'and then you should only try to do that in win-win situations' [CGN, fvj601241__65]

d. en ik denk dat men daarin moet trachten het juiste evenwicht te zoeken.
and I think that one there-in must try the right balance to find
'and I think that one has to find the right balance in that.' [CGN, fvg600012__38]

In (6.45a), *probeert* 'tries' appears in the second pole. The *te*-infintive has an underspecified CLUSTER value, which is not resolved after its combination with *probeert*, as it is an optional clustering verb that does not require its complement to be [CLUSTER +] or [CLUSTER –].

Depending on the types of cluster creepers that are allowed, the constructions in (6.45b–d) can be classified as clustering, non-clustering, or both. In Standard and colloquial Dutch the construction in (6.45b) is ambiguous between clustering and non-clustering, so the CLUSTER value of *probeert wakker te worden* is *boolean*. In this case two structures will be assigned to the construction. The *te*-infintives in (6.45c–d) are in the Nachfeld, i.e. they are contextually resolved as [CLUSTER –] after their combination with the adverbial modifiers in (6.45c) and the NP in (6.45d).

6.5 Word order and branching structure

The examples with verb clusters used in the previous sections of this chapter all contain clusters with the canonical word order, and are analysed as a binary right-branching structure, following a.o. Rentier (1994) and Kathol (2000). The literature study as well as the treebank investigation showed, however, that there is considerable word order variation in Dutch verb clusters.

For Standard Dutch, the treebank investigation showed that the generalizations proposed by Broekhuis & Corver (2015) seem to hold, i.e. that bare infinitives and

te-infinitives follow their selector, while participles can obtain a position to the right or the left of their selecting verb (cf. chapter 1). This is illustrated in (6.46), taken from the treebank investigation in chapter 5.

(6.46) a. want ja je wou er vandaag toch over komen$_1$ praten$_2$?
because yeah you wanted there today actually about come talk
'because you wanted to come and talk about it today, right?' [CGN, fvf600243__110]

b. Ik was erg tevreden dat er in Leuven plaats voor bleek$_1$ te zijn$_2$.
I was very pleased that there in Leuven place for turned out to be
'I was very pleased that there turned out to be a place for it in Leuven.' [dpc-cam-001020-nl-sen.p.12.s.6]

c. Ze vinden het opmerkelijk . . . dat de BBC de zwarte piet krijgt$_1$
they find it remarkable . . . that the BBC the jack of spades gets
toegespeeld$_2$.
passed
'They find it remarkable . . . that the BBC got blamed.' [LASSY, WS-U-E-A-0000000028.p.20.s.3]

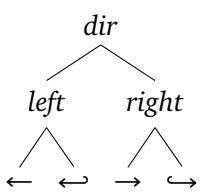
d. . . . terwijl het naburige Lierde de meer bescheiden titel van
. . . while the neighbouring Lierde the more modest title of
mattentaartengemeente toegewezen$_2$ kreeg$_1$.
mattentaarten municipality assigned got
'. . . while the neighbouring town Lierde received the more modest title of 'mattentaarten municipality'.' [LASSY, WR-P-E-I-0000039589.p.11.s.1]

The word order variation within the cluster can be modelled by means of the feature G(O)V(ERN)OR, following Kathol (2000) and Bouma & van Noord (1998). GVOR is a head feature (6.47).[15] Its possible values are presented in (6.48).[16]

[15] In Bouma & van Noord (1998) GVOR is not a head feature. As they employ a flat analysis, the value of GVOR need not be shared with the mother node.

[16] Both Kathol (2000) and Bouma & van Noord (1998) add an additional level to the possible values of GVOR, distinguishing precedence from adjacent precedence, cf. (6.1) (Kathol 2000: 211).

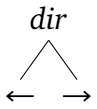
(6.1)



↩, for instance, indicates that the governor has to immediately precede the verbal complement, whereas in the case of ← some other elements can occur in between the verbal complement and

(6.47) *head*: $\left[\text{GVOR} \quad \textit{dir}\right]$

(6.48)



Infinitival complements in Dutch have the feature [GVOR ←], indicating that their governor appears to the left. Past participles have an underspecified value *dir* for GVOR, accounting for the variation in (6.46c–d). The constraints are given in (6.49).

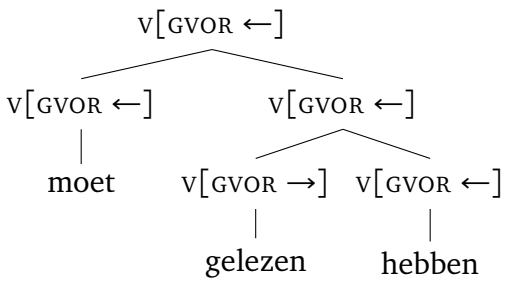
(6.49) a.
$$\left[\text{SS}\,|\,\text{LOC}\,|\,\text{CAT}\,|\,\text{HEAD}\begin{bmatrix}\textit{verb}\\ \text{VFORM} & \textit{inf}\\ \text{GVOR} & \leftarrow\end{bmatrix}\right]$$

b.
$$\left[\text{SS}\,|\,\text{LOC}\,|\,\text{CAT}\,|\,\text{HEAD}\begin{bmatrix}\textit{verb}\\ \text{VFORM} & \textit{psp}\\ \text{GVOR} & \textit{dir}\end{bmatrix}\right]$$

Since GVOR is a head feature, its value is shared between the mother and the head daughter, correctly accounting for the word order in (6.50), as illustrated in (6.51).[17]

(6.50) er zijn toch zo'n paar boeken die ge moet$_1$ gelezen$_3$ hebben$_2$ in uw leven.
there are actually such couple books that you have read have in your life
'actually there are a couple of books that you should have read in your life.' [CGN, fva400503__11]

(6.51)



its selector. As the analysis of verb clusters presented here is binary branching such a distinction is not needed.

[17]Verb-final finite verbs are specified [GVOR ←], as the complementizer is in the first pole; V-initial finite verbs are [GVOR *dir*], as they are the head of the phrase.

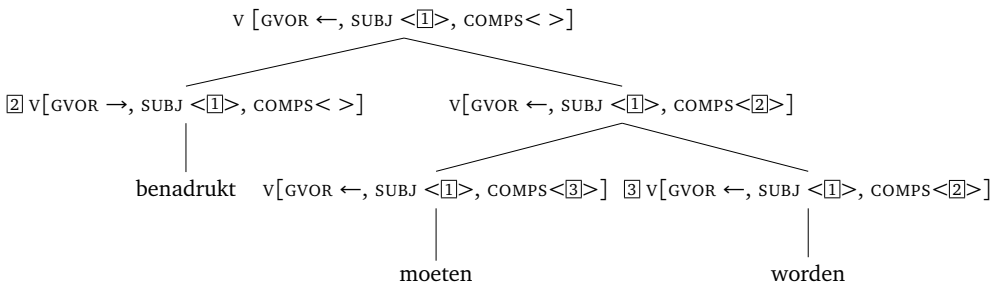
In (6.51), the GVOR value of the past participle is contextually resolved from *dir* to →, allowing for selection on the left of *hebben* 'have'. The verb *hebben* is specified as [GVOR ←]. As GVOR is a head feature, it is shared between *hebben* and the phrase *gelezen hebben*, correctly accounting for the fact that the governing finite verb *moet* 'must' appears to the left.

Constructions in which the participle occurs more to the left in the cluster, as in (6.52), can be dealt with in a similar way.

(6.52) Diversiteit in onze samenleving zou nog veel meer benadrukt$_3$ moeten$_1$
diversity in our society should still much more focused have
worden$_2$.
be
'Diversity in our society should be much more focussed on.' [LASSY, WR-P-P-G-0000000019.p.5.s.14]

As there is no restriction on the type of complements that can be raised, the verb cluster in (6.52) can be analysed as one in which the past participle is raised:

(6.53)

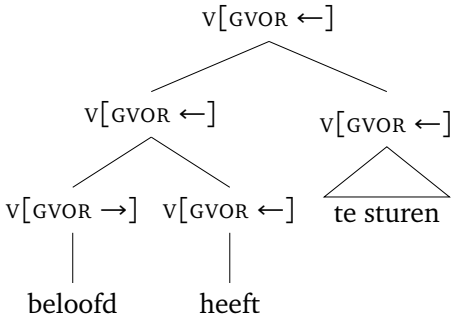


Worden 'be' selects a subject and a participial complement. Its GVOR feature is specified ←, as its selector has to appear to its left. *Moeten* 'must' selects the unsaturated verb *worden*, whose COMPS list is shared with the COMPS list of the phrase *moeten worden*, according to the CRP. Similarly to other raised complements (if any), a raised past participle occurs to the left of its selector. This correctly accounts for the fact that only past participles can be raised (at least in Standard and colloquial Dutch). Since raising involves selection to the left, complement raising of infinitival complements is blocked by the requirement that they appear to the right of their selector, i.e. they have the feature [GVOR ←] in their lexical entry. This blocks the occurrence of verb clusters like **zingen zal willen* 'sing will want'.

Instances of the third construction in which the past participle occurs in front of the auxiliary, can be analysed in a similar way as (6.51). An example is given in (6.54). The tree structure is presented in (6.55).

(6.54) Verder wacht Donner de brief af die premier Godett beloofd heeft te
further wait Donner the letter off that prime minister Godett promised has to
sturen.
send
'Apart from that Donner is waiting for the letter that prime minister Godett has promised to send.'[LASSY, WS-U-E-A-0000000011.p.22.s.15]

(6.55)



The interplay between the GVOR feature and the CRP discussed so far is sufficient to deal with the most common types of word order variation in Dutch verb clusters, i.e. the variation in clusters containing a past participle. The analysis provides a surface-oriented, binary-branching account of such verb clusters, which is different from the accounts provided in Bouma & van Noord (1998) and Kathol (2000).

In some varieties of Dutch,[18] the set of IPP verbs also have an underspecified value for GVOR if they appear as IPP. This can be formulated as constraint on the lexical entries of the IPP selector, i.e. the auxiliaries of the perfect. IPP constructions in Standard Dutch can be modelled by the constraint in (6.56).

(6.56)
$$\left[\begin{array}{l} \textsc{phon} \left\langle \text{hebben} \mid \text{zijn} \right\rangle \\ \textsc{comps} \left\langle \left[\begin{array}{l} \textit{verb} \\ \textsc{head} \left[\begin{array}{l} \textsc{vform}\ \textit{infinitive} \\ \textsc{position}\ \textit{final} \\ \textsc{gvor}\ \leftarrow \end{array}\right] \\ \textsc{lex}\ - \\ \textsc{cluster}\ + \end{array}\right] \right\rangle \end{array}\right]$$

If the verbs of the perfect select a bare infinitive as a complement (i.e. an IPP verb), the complement has to be [CLUSTER +], as IPP verbs are obligatory clustering. In

[18] Mainly Dutch spoken in the provinces of East Flanders and West Flanders in Belgium.

addition, the infinitival complement needs to be phrasal, which is indicated by the feature [LEX –].[19] In this way, two-verb clusters of the type auxiliary-infinitive are excluded (e.g. * *heeft kunnen* 'has can.INF').

In order to deal with the word order in (6.57), which is in fact the Dutch counterpart of the German auxiliary flip, the constraint in (6.56) needs to be adapted to the one in (6.58), in which the GVOR value of the infinitival complement is underspecified.

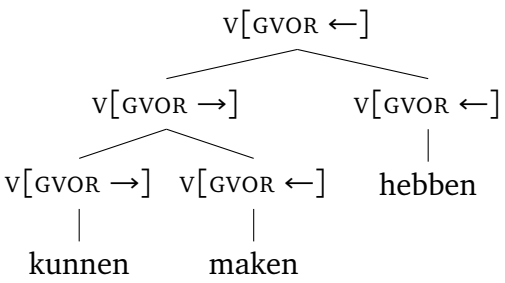
(6.57) a. . . . want ik zou op dat moment geen uh lesvoorbereidingen in uh
. . . because I would on that moment no uh course-preparations in uh
kunnen$_2$ maken$_3$ hebben$_1$ met de PC hè.
can make have with the PC hè
'. . . because I wouldn't have been able to make course preparations with the computer at that moment.' [CGN, fvb400165__167]

b. . . . terwijl dat 'k ik naar buiten gaan$_2$ kijken$_3$ ben$_1$.
. . . while that I I to outside go look am
'. . . while I was going to look outside.' [CGN, fva400388__48]

(6.58)
$$\left[\begin{array}{l}\textsc{phon}\ \langle \text{hebben} \mid \text{zijn} \rangle \\ \textsc{comps}\ \left\langle \left[\begin{array}{l} \textit{verb} \\ \textsc{head}\ \left[\begin{array}{l}\textsc{vform}\ \textit{infinitive} \\ \textsc{position}\ \textit{final} \\ \textsc{gvor}\ \textit{dir}\end{array}\right] \\ \textsc{lex}\ - \\ \textsc{cluster}\ +\end{array}\right] \right\rangle\end{array}\right]$$

In (6.57a) the GVOR value is underspecified in the lexical entry of the IPP *kunnen* 'can', but the tree representation in (6.59) illustrates how it is contextually resolved.

(6.59)



[19] The only purpose of the LEX feature in this analysis is to indicate whether a complement needs to be lexical ([LEX +]) or phrasal ([LEX –]).

Also in this case, the sharing of GVOR between the head daughter *kunnen* and the mother node correctly accounts for the word order variation in IPP constructions. Note that the GVOR value is only underspecified for the IPP forms, in order to exclude the erroneous 2-3-1 order in e.g. * *kunnen maken zal* 'can make will'.

A final type of word order variation observed in the treebanks concerns two-verb clusters containing a finite verb and an infinitive, as in (6.60).

(6.60) a. . . . een apart verkeersreglement voor Vlaanderen en Wallonië dat is
. . . a separate traffic-regulations for Flanders and Wallonia that is
wat ons redden$_2$ gaat$_1$.
what us save will
'. . . separate traffic regulations for Flanders and Wallonia is what will save us.' [CGN, fvl600253__22]

b. . . . om ervoor te zorgen dat dit nooit meer gebeuren$_2$ zal$_1$.
. . . to there-for to make-sure that this never again happen will
'. . . in order to make sure that this will never happen again.' [LASSY, dpc-vhs-000725-nl-sen.p.7.s.4]

c. Hardy was waar hij wezen$_2$ wilde$_1$.
Hardy was where he be wanted
'Hardy was where he wanted to be.' [CGN, fvo800575__25]

The literature is not entirely conclusive for which verbs this variation is allowed, see for instance Broekhuis & Corver (2015: 1095–1097) for a discussion. Most authors mention the core modals, as well as the aspectual *gaan* 'go' and the subject control verb *willen* 'want' as verbs that can appear to the right of their selector if they are finite.

The set of verbs that allow this type of inversion can be described by means of the constraint in (6.61):[20]

[20]Kathol (2000: 200) mentions constructions like (6.60) and concludes that in Dutch, non-finite verbs are all specified [GVOR *dir*]. As bare infinitives can only precede their selector in a limited set of constructions, this statement is too general.

(6.61)
$$\left[\begin{array}{l} \textsc{head} \left[\begin{array}{l} \textsc{vform}\ \textit{finite} \\ \textsc{position}\ \textit{final} \end{array}\right] \\ \textsc{comps} \left\langle \left[\begin{array}{l} \textit{verb} \\ \textsc{head} \left[\begin{array}{l} \textsc{vform}\ \textit{infinitive} \\ \textsc{position}\ \textit{final} \\ \textsc{gvor}\ \textit{dir} \end{array}\right] \\ \textsc{lex}\ + \end{array}\right] \right\rangle \end{array}\right]$$

While infinitives are canonically specified as [GVOR ←], the constraint in (6.61) leaves the value of GVOR of the infinitival complement underspecified, allowing for selection to the left or to the right if it is selected by a finite clustering verb that can occur in this type of construction. As this type of inversion is only allowed in two-verb clusters, the infinitival complement must be lexical, which is indicated by LEX +. This correctly rules out clusters like **willen proberen zal* 'want try will' or **kunnen zwemmen moet* 'can swim must'.

While the GVOR feature is used by Bouma & van Noord (1998) and Kathol (2000) in order to account for word order variation within the cluster, it can also be employed to model word order variation between heads and their dependents in general, as it is not restricted to a particular part-of-speech, cf. (6.47). For instance, NPs and APs, and particles are specified as [GVOR →] if their head is defined as [POSITION *final*]. This ensures that cluster creepers appear to the left of the selecting verb, so constructions of the type **zal bellen op* 'will call up' are ruled out. PPs have an underspecified value for GVOR, as they can appear before and after their head, cf. the PP-over-V constructions that will be discussed in chapter 7. Note that GVOR only indicates the direction of selection with respect to the head; the word order of the non-verbal complements with respect to each other is indicated by their order on the COMPS list of the verb, as explained in chapter 3 (section 3.2.2).

In sum, this section presented how the most common types of word order variation in Dutch verb clusters can be accounted for. Following Bouma & van Noord (1998) and Kathol (2000), the GVOR feature is used to model this. While the lexical specification of this feature and the application of the CRP is sufficient to model the variation in verb clusters containing a participle, additional constraints are needed in order to model word order variation in IPP constructions and in a limited set of two-verb clusters containing a finite verb and a bare infinitive. In addition, the GVOR

feature can be used to model word order between heads and their dependents in general, as it is not restricted to verbal signs only.

6.6 Conclusion

In this chapter it was argued that the *argument inheritance* or *generalized raising* treatment discussed in chapter 3 poses a number of problems if it is applied to Dutch. The main problem lies in the fact that its interaction with the binding principles and the passive lexical rule yields erroneous predictions.

An alternative is proposed, employing different devices for subject raising and complement raising: While the former is modelled in terms of lexical constraints, as in English, the latter is modelled in terms of a constraint on headed phrases. In order to avoid overgeneration, a constraint is added to prevent complement raising out of CPs and verb-initial VPs. In addition, it was shown how cluster creeping and word order variation within Dutch verb clusters can be accounted for.

The data discussed here were mainly from Dutch, but as pointed out by Pollard (1994) the argument inheritance approach also poses problems with respect to the binding principles in German, indicating that the analysis proposed in this chapter might be applicable to German as well. However, the occurrence of the remote passive in German might be a reason to employ a uniform argument attraction model via the selecting verbs. Still, the fact that not all speakers of German allow such constructions suggests that the argument inheritance mechanism in the Hinrichs-Nakazawa style is not suitable for all varieties of German. The investigation to what extent the complement raising principle is applicable to other languages than Dutch is left for future work.

CHAPTER 7

Beyond verb clusters

So far, the focus of the discussion was on raising out of non-finite verbal complements. This, however, is not the only type of raising that the complement raising principle (CRP) in (6.16) allows. It also allows raising out of non-verbal phrases, such as adjectival and prepositional phrases. This will be illustrated in section 7.1. In section 7.2 the case of complement raising out of prepositional complements will be discussed in more detail, as those constructions are canonically dealt with as instances of complement extraction. It will be motivated that there are instances of raising out of PPs that are better analysed as instances of complement raising. Besides the question how cases of complement raising can be differentiated from cases of complement extraction, the constraints on both complement raising and extraction out of PPs will be dealt with. Section 7.3 concludes this chapter.

7.1 Complement raising out of non-verbal phrases

While the CRP is originally formulated to analyse raising out of verbal complements, it can be applied to deal with raising out of non-verbal complements (section 7.1.1). Furthermore, the CRP can be used to analyse raising out of subjects and adjuncts as well (section 7.1.2).[1]

7.1.1 Complement raising out of non-verbal complements

Some examples of complement raising out of adjectival phrases are given in (7.1).

[1] This section is based on Van Eynde & Augustinus (2013, 2014).

(7.1) a. . . . dat we *die hittegolf* nog steeds niet **kwijt** zijn!
. . . that we that heat wave still always not lost are
'. . . that we are not finished with that heat wave yet!' [LASSY, WS-U-E-A-0000000221.p.32.s.2]

b. . . . dat de bevolking van Zimbabwe *haar huisbakken dictator* onderhand
. . . that the people of Zimbabwe her home-grown dictator by now
meer dan **beu** is.
more than fed-up is
'. . . that the people of Zimabwe are more than fed up with their homegrown dictator by now.' [LASSY, WR-P-P-I-0000000219.p.4.s.4]

The italicized nominals are complements of the predicative adjectives in bold, but they are not realized within the AP. Instead, they are raised and realized in the left part of the Mittelfeld, preceding the VP adjuncts.

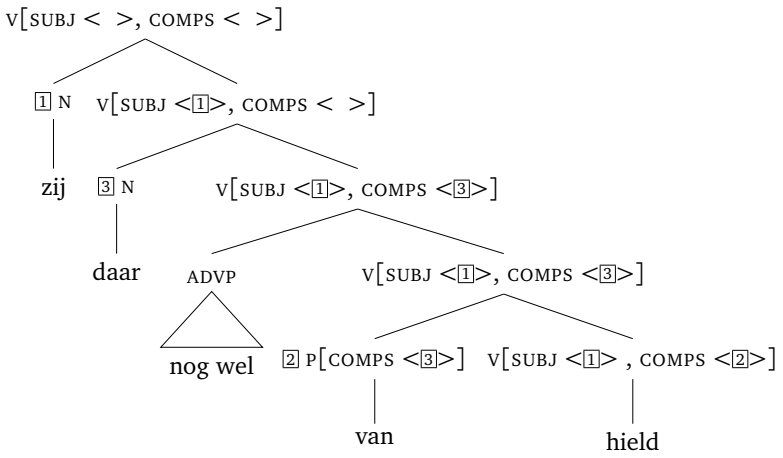
Complement raising also subsumes the instances of adposition stranding in (7.2).[2]

(7.2) a. . . . dat zij *daar* nog wel **van** hield.
. . . that she that+R still rather of liked
'. . . that she rather liked it.' [CGN, fna000741__12]

b. . . . als je *er* pas achteraf **over** nadenkt, is het misschien te laat.
. . . if you it+R only later about think-of, is it maybe too late
'. . . if you only think about it afterwards, it may be too late.' [LASSY, WR-P-P-C-0000000047.txt-10]

Also here, the italicized pronouns are complements of the adpositions in bold, but they are not realized within the PP. Instead, they are raised and realized in the left part of the Mittelfeld as well, illustrated by the representation of (7.2a) in (7.3).

(7.3)



[2]Adposition stranding will be discussed more in depth in section 7.2.

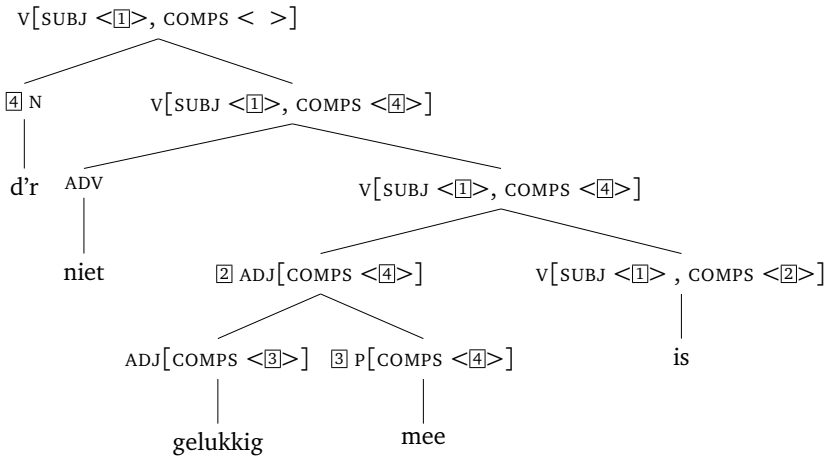
The unsaturated COMPS requirement of the adposition (3) is inherited by the verbal projection and discharged after the addition of *daar* 'that+R'.

Since there are adjectives which take adpositional complements, complement raising can be applied iteratively, as in (7.4).

(7.4) dat die *d'r* niet **gelukkig** mee is.
that that one that+R.RED not happy with is
'that she is not happy with that. [CGN, fva400731__123]'

The pronominal complement *d'r* 'there' is first raised out of the PP, then out of the predicative AP, and finally out of the verb-final VP, as illustrated in (7.5).

(7.5)



The adjective's requirement for an adpositional complement (3) is immediately saturated, and so is the verb's requirement for a predicative complement (2), but the adposition's requirement for a nominal complement (4) is not. It is appended to the COMPS list of the mother and propagated up the tree, till the point where the addition of *d'r* triggers its cancellation. Notice that the requirement for a nominal complement (4) figures in the COMPS lists of the adposition and the nodes which dominate it, but not in the COMPS lists of the adjective or the verb.

The example in (7.4) is comparable to the iterative subject raising in sequences like (7.6).

(7.6) He does not seem to be likely to win this game.

The surface subject of *does* is the understood subject of *win this game*, and the relation is mediated by a sequence of subject raising lexemes, including *to*, *likely*, *be*, *seem* and *does*.

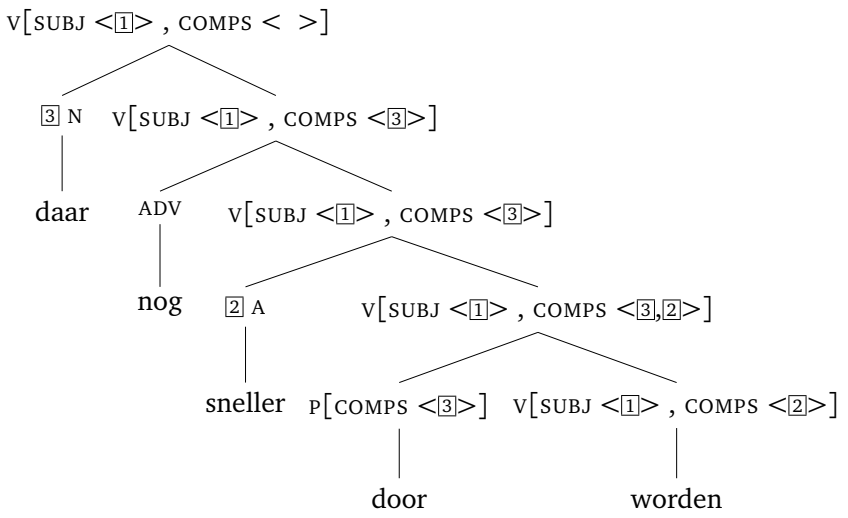
7.1.2 Complement raising out of subjects and adjuncts

The previous examples all concerned raising out of complements, but the CRP does not require this: It also allows the non-head daughter to be a functor or a subject. (7.7), for instance, is an example of complement raising out of a PP adjunct.

(7.7) heeft uh overigens z'n haar uh wat korter geknipt. zal *daar*
has uh for the rest his.RED hair uh somewhat shorter cut.PSP will.FIN that+R
nog sneller **door** worden. wie weet.
still faster by become.INF who knows
'Apart from that (he) has cut his hair uh somewhat shorter. Will become even faster because of that. Who knows.' [CGN, fni007447__27]

The *door*-phrase in this sentence is not a complement of the verb, but an adjunct. It specifies the cause of the swiftness. Its COMPS requirement is not immediately saturated, but propagated in the by now familiar way.

(7.8)



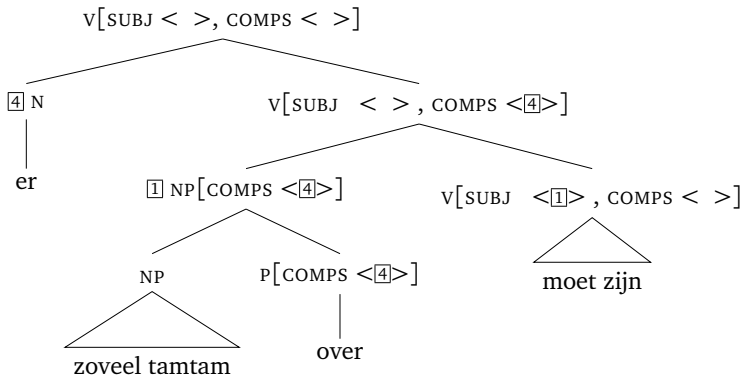
In the generalized raising treatment, Hinrichs-Nakazawa style, this would require special measures: Since adjuncts are not selected by their head sister, the latter cannot inherit the former's COMPS requirement in the usual way. No such complications are needed in the treatment of complement raising.

Raising out of subjects is exemplified in (7.9).

(7.9) ...dat *er* zoveel tamtam **over** moet zijn voordat dat gewoon naar buiten
...that it+R so much fuss about must be before it just to outside
komt.
comes
'...that there should be so much fuss about it before it becomes public.' [CGN, fnf007187__41]

The italicized pronoun is a complement of the adposition *over* 'about', which heads the PP adjunct of the NP *zoveel tamtam* 'so much fuss' which in its turn heads the subject of the clause, as spelled out in (7.10).

(7.10)



7.1.3 Conclusion

In sum, this section has shown how the CRP, originally formulated to treat raising out of verbal complements, easily extends to raising out of non-verbal phrases. In section 7.1.1 it was shown that it can also account for raising out of non-verbal complements, since it does not put any constraints on the syntactic category of the non-head daughter. Furthermore, it allows raising out of subjects and adjuncts, since the CRP applies to all headed phrases, as was shown in section 7.1.2.

7.2 Adposition stranding

As the treebank examples in section 7.1 already indicated, Dutch adpositions can be stranded, typically if their complement is an R-pronoun. The complement usually

appears in the left part of the Mittelfeld or in the Vorfeld. In HPSG this is canonically modeled in terms of extraction, making use of non-local devices such as SLASH and BIND. This section will argue that the extraction analysis is indeed appropriate for cases in which the complement is realized in the Vorfeld, but it proposes a complement raising analysis for the cases in which the complement is realized in the Mittelfeld. The motivation for this distinction is that extraction to the Vorfeld is a long-distance dependency, whereas complement raising in the Mittelfeld involves a medium-distance dependency. Some examples of complement raising out of PPs were already given in section 7.1, but this section discusses more in depth in which cases complement raising can be applied, and when it should be blocked.[3]

7.2.1 Adposition stranding in Dutch

In Dutch, adpositions canonically precede their complement.

(7.11) a. Ze zegt dat ze soms nog [aan hem/Hans] denkt.
she says that she sometimes still of him/Hans thinks
'She says that she still thinks of him/Hans from time to time.'

b. * Ze zegt dat ze soms nog [hem/Hans aan] denkt.
she says that she sometimes still him/Hans of thinks

However, if the complement is a demonstrative pronoun, such as *dit* 'this' or *dat* 'that', it takes another form, the so-called R-form, and precedes the adposition.[4]

(7.12) a. * Ze zegt dat ze soms nog [aan dat/dit] denkt.
she says that she sometimes still of that/this thinks

b. Ze zegt dat ze soms nog [daar/hier aan] denkt.
she says that she sometimes still that+R/this+R of thinks
'She says that she still thinks of that/this from time to time.'

The same holds for the personal (neuter) *het* 'it' and the interrogative/relative *wat* 'what', which alternate with *er* 'it+R' and *waar* 'what+R' respectively. The alternation also applies to the quantifying *iets* 'something', *niets* 'nothing' and *alles* 'everything', but for these pronouns it is optional: (7.13a) and (7.13b) are equally well-formed.[5]

[3] A previous version of this section appeared in Van Eynde & Augustinus (2014).

[4] The sequence of the R-pronoun and the adposition is often treated as an orthographic unit, as in *daaraan* and *hieraan*.

[5] The sequence of *(n)ergens/overal* and an adposition is not treated as a single unit in the orthography.

(7.13) a. Ze zegt dat ze soms gewoon [aan niets] denkt.
she says that she sometimes simply of nothing thinks
'She says that she simply thinks of nothing from time to time.'

b. Ze zegt dat ze soms gewoon [nergens aan] denkt.
she says that she sometimes simply nothing+R of thinks
'She says that she simply thinks of nothing from time to time.'

Table 7.1 provides a survey of the pronouns which show the [–/+ R] alternation. What they have in common is that they all canonically (but not exclusively) denote a thing rather than a person (Broekhuis 2013: 293–294).

Pronoun	[–R]	[+R]	
Personal	*het*	*er*	'it'
Demonstrative	*dat*	*daar, d'r*	'that'
	dit	*hier*	'this'
Interrogative/Relative	*wat*	*waar*	'what'
Quantifying	*iets*	*ergens*	'something'
	niets	*nergens*	'nothing'
	alles	*overal*	'everything'

Table 7.1: The Dutch pronouns with an R-form

A peculiar property of the R-pronouns is that they tend to be realized outside of the PP: They typically end up in the left part of the Mittelfeld, preceding the VP adjuncts, as in (7.14), or in the Vorfeld, as in (7.15).

(7.14) a. Ze zegt dat ze *daar* soms nog [__ aan] denkt.
she says that she that+R sometimes still of thinks
'She says that she still thinks of it from time to time.'

b. We hebben *er* toen een lied [__ over] gezongen.
we have it+R then a song about sung
'We have sung a song about it.'

(7.15) a. *Waar* denk je dat ze [__ op] wachten?
what+R think you that they for wait
'What do you think they are waiting for?'

b. *Hier* kunnen we echt niet [__ op] wachten.
this+R can we really not for wait
'This we really cannot wait for.'

The result of this non-local realization is that the adposition is left alone: It is stranded in the right part of the Mittelfeld. The phenomenon has been studied extensively. Descriptive surveys are provided in Haeseryn et al. (1997) and Broekhuis (2013), transformational treatments in Van Riemsdijk (1978) and Bennis (1986), and HPSG treatments in Rentier (1993) and Bouma (2000).[6] Both of the latter treat the phenomenon in terms of extraction, employing non-local devices such as SLASH and BIND.

7.2.2 No complement raising out of P-initial PPs

In sections 7.1.1 and 7.1.2, it was already shown how the complement raising device used to model scrambling in the Mittelfeld can be used to model the raising out of PPs as well, no matter whether those PPs are complements of verbs, complements of other categories, or adjuncts. It is, hence, a very powerful device. As is the case for raising out of verbal complements, there are restrictions on raising out of PPs.

A general constraint on Dutch adposition stranding is that it only affects complements which precede the adposition. Complements which follow the adposition must be realized within the PP.

(7.16) a. Ze zegt dat ze soms nog [aan hem/Hans] denkt.
she says that she sometimes still of him/Hans thinks
'She says that she still thinks of him/Hans from time to time.'

b. * Ze zegt dat ze *hem/Hans* soms nog [aan __] denkt.
she says that she him/Hans sometimes still of thinks

This is confirmed by the locative adverbs. They are homophonous to the R-pronouns, but in contrast to the latter they follow the adposition in PPs and must be realized within the PP.

(7.17) a. ... dat ze volgens ons niet [van hier] zijn.
... that they according-to us not from here are
'... that they are not from here according to us.'

b. * ... dat ze volgens ons niet [hier van] zijn.
... that they according-to us not here from are

c. * ... dat ze *hier* volgens ons niet [van __] zijn.
... that they here according-to us not from are

[6] The phenomenon also occurs in German, albeit on a smaller scale. See Fleischer (2002) and Duden (2006) for a descriptive overview, and Müller (1995) for an HPSG analysis.

The relevance of the linear order is also clear from the contrast between (7.18a) and (7.18b).

(7.18) a. ... dat de auto dagenlang [in de garage] stond.
... that the car days-long in the garage stood
'... that the car stood in the garage for days.'

b. ... dat ze achteruit [de garage in] reed.
... that she backward the garage in drove
'... that she drove backward into the garage.'

The prepositional PP in (7.18a) requires in situ realization of its complement, but its postpositional counterpart in (7.18b) allows raising.[7]

(7.19) a. * ... dat de auto *de garage* dagenlang [in __] stond.
... that the car the garage days-long in stood

b. ... dat ze *de garage* achteruit [__ in] reed.
... that she the garage backward in drove
'... that she drove backward into the garage.'

This suggests that P-initial PPs are islands for complement raising. This is confirmed by the contrast between (7.20) and (7.21).

(7.20) a. Heb jij al [een boek [daar over]] gelezen?
have you already a book that+R about read?
'Have you already read a book about that?'

b. Heb jij *daar* al [een boek [__ over]] gelezen?
have you that+R already a book about read?
'Have you already read a book about that?'

(7.21) a. Heb jij al [aan een boek [daar over]] meegewerkt?
have you already on a book that+R about collaborated?
'Have you already collaborated on a book about that?'

b. * Heb jij *daar* al [aan een boek [__ over]] meegewerkt?
have you that+R already on a book about collaborated?

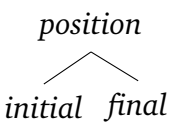
[7]The distinction corresponds to a difference in interpretation: While the prepositional PP has a locational interpretation, the postpositional one has a directional interpretation. Smessaert et al. (2014) point out that spatial descriptions with a postpositional phrase always have a dynamic interpretation, expressing destination with attained goal. Prepositions are typically static, cf. (7.18a). In combination with a motion verb, however, they are sometimes ambiguous between static (locational, motion-in-place) and dynamic (translocation, directional), e.g. *in het water springen* 'jump in/into the water' (Smessaert et al. 2014: 130–131).

Both the in situ and the raised construction in (7.20) are well-formed, while in (7.21) only the in situ construction is well-formed: The addition of the preposition *aan* 'on' blocks the raising of the complement.

In sum, complements can be raised out of a P-final PP, but not out of a P-initial PP. Besides, the raised complement cannot only be an R-pronoun, but also a full NP, as in (7.19b).

In chapter 6 it was shown that complement raising out of verb-final VPs is possible, while complement raising out of verb-initial VPs and (C-initial) CPs is blocked (see sections 6.2.2 and 6.3.2). It is possible to block complement raising out of P-initial PPs in a similar fashion, i.e. by means of the POSITION feature. Its type declaration and the inventory of its values were already given in (6.32) and (6.33), and repeated in (7.22) and (7.23) respectively.

(7.22) *non-nominal*: [POSITION *position*]

(7.23)


Similar to the verbs, the Dutch adpositions come in three types. Some are inherently initial, such as *met* 'with', *tot* 'to, till', *te* 'at, to' and *sinds* 'since, for', some are inherently final, such as *mee* 'with', *toe* 'to, till', *af* 'from' and *heen* 'towards', and some are used either way, such as *in* 'in', *op* 'up, on', *aan* 'on' and *van* 'of'. Table 7.2 provides a survey, including the classification of the verbs and complementizers.

POSITION	Adpositions	Verbs	Complementizers
Initial	*met, tot, te, sinds*	imperative	*dat, of, als, dan, om*
Final	*mee, toe, af, heen*	non-finite	–
Underspecified	*in, op, aan, van*	non-imperative finite	–

Table 7.2: The POSITION values of adpositions, verbs, and complementizers

Assuming that the underspecified values are resolved contextually, the constraint which blocks complement raising (6.34), repeated in (7.24), is also applicable to P-initial adpositions:

(7.24) $\left[\begin{array}{l} \textit{hd-ph} \\ \textsc{synsem}\,|\,\textsc{loc}\,|\,\textsc{cat}\,|\,\textsc{head}\,|\,\textsc{position} \quad \textit{initial} \end{array}\right] \Rightarrow \left[\textsc{synsem}\,|\,\textsc{loc}\,|\,\textsc{cat}\,|\,\textsc{comps} \quad \langle\ \rangle\right]$

What (7.24) says, is that phrases whose POSITION value is resolved to *initial* must have an empty COMPS list. From this it follows that complements cannot be raised out of V-initial VPs, nor out of C-initial CPs, nor out of P-initial PPs. Technically, the restriction to verbs, complementizers and adpositions is due the fact that only these have the POSITION feature. Empirically, it is motivated by the fact that the other lexical categories do not abide by the constraint. APs and NPs, for instance, allow complement raising, also if the head precedes its dependents. This was shown for the R-pronouns in (7.4) and (7.9), but it also holds for other kinds of dependents of adjectives and nouns, such as the italicized PPs in (7.25b) and (7.26b).

(7.25) a. Ze zullen volgens mij niet [blij met die resultaten] zijn.
they will according-to me not happy with those results be

b. Ze zullen *met die resultaten* volgens mij niet [blij __] zijn.
they will with those results according-to me not happy be
'They will not be happy with those results according to me.'

(7.26) a. Ze hebben om het uur [foto's van elk dier] gemaakt.
they have about the hour pictures of each animal made

b. Ze hebben *van elk dier* om het uur [foto's __] gemaakt.
they have of each animal about the hour pictures made
'Every hour they made pictures of each animal.'

The restriction to verbs, complementizers and adpositions is, hence, justified.

Finally notice that the distinction between P-initial and P-final PPs is not only relevant to treat the linear order within the PP and to avoid complement raising out of P-initial PPs. It furthermore correlates with another phenomenon: PP-over-V, i.e. the realization of a PP in the Nachfeld. It is possible for P-initial PPs, as illustrated in (7.27), but not for P-final PPs, as shown in (7.28–7.29).

(7.27) a. ... dat we nog steeds [**op** een goede afloop] hopen.
... that we still always for a good outcome hope
'... that we are still hoping for a good outcome.'

b. ... dat we nog steeds hopen [**op** een goede afloop].
... that we still always hope for a good outcome
'... that we are still hoping for a good outcome.'

(7.28) a. ... dat we nog steeds [daar **op**] hopen.
... that we still always that+R for hope
'... that we are still hoping for that.'

b. * ... dat we nog steeds hopen [daar **op**].
... that we still always hope that+R for

(7.29) a. ... dat hij voorzichtig [de garage **in**] reed.
... that he carefully the garage in drove
'... that he drove carefully into the garage.'

b. * ... dat hij voorzichtig reed [de garage **in**].
... that he carefully drove the garage in

When the constraint in (7.24) is combined with the observations about PP-over-V, it correctly accounts for the fact that adpositions cannot be stranded in the Nachfeld.

(7.30) a. * ... dat we *een goede afloop* nog steeds hopen **op**
... that we a good outcome still always hope for

b. * ... dat hij *daar* nog steeds hoopt **op**
... that he that+R still always hopes for

c. * ... dat hij *de garage* voorzichtig reed **in**
... that he the garage carefully drove in

(7.30a) is ill-formed, since (7.24) does not allow to raise a complement out of a P-initial PP, and (7.30b–7.30c) are illformed, since P-final PPs are not allowed in the Nachfeld.

7.2.3 Complement raising vs complement extraction out of PPs

As shown in section 6.2.2, complement raising out of verbal complements should be differentiated from complement extraction. As for extraction out of PPs, it is clear that P-final PPs allow it, see (7.15). Some other examples are those in (7.31).

(7.31) a. *Welke garage* denk je dat hij toen achteruit [__ in] reed?
which garage think you that he then backward in drove?
'Which garage do you think he drove into backward?'

b. *Daar* hebben we toen met de baas [__ over] gesproken.
that+R have we then with the boss about spoken
'That we talked about with the boss then.'

P-initial PPs, by contrast, are islands for extraction.

(7.32) a. * *Welke garage* denk je dat ze [in __] liggen?
which garage think you that they in lie?

b. * *Daar* hebben we toen [met de baas [__ van]] gesproken.
that+R have we then with the boss of talked

As indicated by the bracketing, the stranded *van* 'of' in (7.32b) is a PP-adjunct of *baas* and, hence, included in the PP that is introduced by *met* 'with'. It thus contrasts with the stranded *over* 'about' in (7.31b), which is a PP-complement of the verb *gesproken* 'spoken'.

In contrast to the P-initial PPs, C-initial CPs and V-initial VPs allow complement extraction. This was shown in (6.22) and (6.24), repeated in (7.33) and (7.34).

(7.33) *Wie* [beweert ze [dat ze in Parijs [__] ontmoet hebben]]?
who claims she that they in Paris met have
'Who does she claim they met in Paris?'

(7.34) *Dat boek* zal Peter [__] toch niet kunnen vinden.
that book shall Peter anyway not can find
'That book, Peter will not be able to find anyway.'

The constraint on complement extraction is, hence, less restrictive than the one on complement raising:

(7.35) $$\left[\begin{array}{l} \textit{hd-ph} \\ \\ \textsc{synsem}\,|\,\textsc{loc}\,|\,\textsc{cat}\,|\,\textsc{head} \left[\begin{array}{ll} \textit{adposition} & \\ \textsc{position} & \textit{initial} \end{array}\right] \end{array}\right] \Rightarrow \left[\textsc{synsem}\,|\,\textsc{nonloc}\,|\,\textsc{slash}\ \{\ \}\right]$$

In plain words, P-initial PPs must have an empty SLASH set, but this constraint does not extend to V-initial VPs nor to clauses which are introduced by a complementizer.

7.2.4 A comparison with the uniform extraction analysis

A distinctive property of the analysis of adposition stranding presented in this section is that it is treated as the result of either complement raising or complement extraction. In this respect it differs from the existing HPSG treatments which see it as the result of complement extraction only. Rentier (1993) and Müller (1995) take the uniform extraction analysis for granted and focus mainly on the issue of how it can be spelled out in formal detail. Bouma (2000), by contrast, considers argument inheritance as an alternative for the uniform extraction analysis, but then argues against it. His four arguments will be discussed in this section.

Raising versus extraction

> "Prepositions which do not allow extraction (such as *met*) cannot be associated with an R-pronoun in the Mittelfeld either. If two different mechanisms are used to account for these two phenomena, such generalizations are easily lost." (Bouma 2000: 69)

The answer to this objection is threefold. First, it is true that separate constraints are needed to account for complement raising and complement extraction, see (7.24) and (7.35) respectively. This, however, is motivated by the fact that the former also subsumes the verbs and the complementizers, while the latter does not. By using a single constraint the generalization is lost that the constraint on complement raising also subsumes V-initial VPs and C-initial CPs.

Second, the empirical argument for differentiating raising from extraction is also valid for the R-pronouns. The reduced forms *er* and *d'r* can be raised, as in (7.36a), but they cannot be extracted, as shown in (7.36b).

(7.36) a. We hebben *daar/er/d'r* een liedje [__ over] gezongen.
we have that+R/it+R a song about sung
'We sang a song about that.'

b. *Daar/*er/*d'r* hebben we een liedje [__ over] gezongen.
that+R/*it+R have we a song about sung
'That we sang a song about.'

Third, there are languages, such as English, which allow adposition stranding as a result of extraction, but not as a result of raising.

(7.37) a. *What* did you say she sang a song [about __]?

b. *That man* I never want to talk [to __] again.

(7.38) a. * I once heard *it* a song [about __].

b. * You should never *that* talk [about __] again.

The ban on complement raising follows from the Empty COMPS Constraint (ECC) as defined in Ginzburg & Sag (2000), see (6.27), repeated in (7.39).

(7.39) $$\textit{phrase} \quad \Rightarrow \quad \left[\text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{COMPS} \quad \left\langle \right\rangle \right]$$

This constraint requires all phrases to have an empty COMPS list and is, hence, much more restrictive than (7.24), which requires this only for V-initial VPs, P-initial PPs, and C-initial CPs.[8]

In sum, the use of separate constraints on complement raising and complement extraction is motivated by the fact that they have a different range of application, as well as by the fact that there are languages which have one but not the other.

PP-internal order

> "As argument inheritance normally involves the composition of two COMPS lists, R-pronouns would have to be allowed on COMPS, even though they can, apart from a few exceptional cases, never appear in a position following the preposition." (Bouma 2000: 69)

This objection is based on the assumption that a nominal can only be a complement of an adposition if it follows that adposition, as in (7.11a) and (7.13a). This assumption, though, is hardly tenable in view of the fact that R-pronouns canonically precede the adposition when they are realized within the PP, as in (7.12b) and (7.13b), repeated in (7.40).

(7.40) a. Ze zegt dat ze soms nog [daar aan] denkt.
she says that she sometimes still that+R on thinks
'She says that she still thinks about it from time to time.'

b. Ze zegt dat ze soms gewoon [nergens aan] denkt.
she says that she sometimes simply nothing+R of thinks
'She says that she simply thinks of nothing from time to time.'

It is also contradicted by the PP-internal order in (7.18b), repeated in (7.41).

(7.41) ... dat ze achteruit [de garage in] reed.
... that she backward the garage in drove
'... that she drove backward into the garage.'

Rentier (1993: 116), who just like Bouma assumes that Dutch PPs must be *pre*positional, mentions (7.41) as a possible counterexample for his claim that Dutch has no postpositions, but then casts doubt on the adpositional status of *in*, claiming that it might be a particle. Notice though, that the adposition in (7.41) is clearly distinct

[8] It might make sense to restrict the ECC to *headed* phrases, since coordinate phrases may consist of unsaturated words, as in *he buys and sells cars* and *are you for or against the war on terror*.

from the separable verb particle *in* of *inrijden*, a transitive verb denoting the activity of preparing a vehicle (car, bike, bus, ...) for use on the road. For detailed argumentation that postpositions like the one in (7.41) are distinct from particles, see Van Riemsdijk (1978: 90–108).

In addition, given that Dutch has V-final VPs and A-final APs, as shown in (7.42), the existence of P-final PPs is just what one expects.

(7.42) ... dat hij [haar fratsen beu] is.
... that he her antics fed-up is
'... that he is fed up with her antics'.

In fact, Dutch is widely assumed to be predominantly head-final.

Argument inheritance

> "The set of argument inheritance verbs must now not only contain auxiliaries and modals, but all verbs which select a (prepositional) complement. Examples such as *Kim is er tevreden mee* introduce further complications for an argument inheritance approach, as it suggests that predicative adjectives and nouns must be argument inheritors as well." (Bouma 2000: 69)

This is a problem indeed for the generalized raising treatment but not for complement raising. In fact, it is one of the reasons why the CRP is modelled in terms of a constraint on headed phrases rather than in terms of a lexical constraint. The latter is only used to model subject raising and is, hence, limited to auxiliaries, modals and a few other verbs. It is not necessary to extend this to all the verbs, adjectives and nouns which select a PP complement, since the unsaturated COMPS requirements are propagated directly from the non-head daughter to the mother, see (7.3), (7.5) and (7.8).

Amalgamation of syntactic functions

> "In an argument inheritance approach, the relationship between valence and syntactically realized arguments has to be one-on-one, and thus there is no room for amalgamation of syntactic functions." (Bouma 2000: 69)

This objection requires a more lengthy rebuttal. To see what is meant with amalgamation, notice that *er* and *d'r* are not only used as R-pronouns and locative adverbs,

but also as the semantically vacuous subject of existential clauses and impersonal passives, as in (7.43).

(7.43) a. Er/d'r staat een artikel over die mislukte aanslag in de krant.
there stands an article about that failed coup in the newspaper
'There is an article about that failed coup in the newspaper.'

b. Er/d'r wordt nog elke dag over die mislukte aanslag geschreven.
there is still every day about that failed coup written
'That failed coup is still written about every day.'

These uses of *er* and *d'r* can be seen as the nominative counterparts of the non-nominative R-pronouns in PPs. If a clause contains both a nominative and a non-nominative R-pronoun, there is a tendency to drop the latter.

(7.44) a. Er/d'r staat (er) een artikel [__ over] in de krant.
there stands (it+R) an article about in the newspaper
'There is an article about it in the newspaper.'

b. Er/d'r wordt (er) nog elke dag een artikel [__ over] geschreven.
there is (it+R) still every day an article about written
'Every day an article is written about it.'

If the two occurrences are adjacent, the elision is even obligatory.

(7.45) a. ... dat er/d'r (*er) een artikel [__ over] in de krant staat.
... that there (*it+R) an article about in the newspaper stands
'... that there is an article about it in the newspaper.'

b. ... dat er/d'r (*er) een artikel [__ over] geschreven wordt.
... that there (*it+R) an article about written is
'... that there is an article written about it.'

Bouma (2000: 73) treats the clauses with a single occurrence of *er/d'r* as instances of function amalgamation: He assumes that the pronoun simultaneously fulfills two functions in such clauses.[9] This amalgamation, he claims, is impossible to model in terms of argument inheritance, since that device does not allow for discrepancies between valence and syntactically realized arguments.

In the analysis presented here, there is no function amalgamation. Instead, the first *er* tokens in (7.44–7.45) have only one function, i.e. subject of the verb. The

[9]Technically, the amalgamation is modelled in terms of structure sharing: The LOCAL value of the subject is identified with the SLASH value of the adposition as well as with the BIND value of the verb.

homophonous raised pronouns are not identified with that subject, but simply omitted.[10] Independent evidence for this analysis is provided by the fact that the omission also occurs in clauses which do not contain another instance of *er*, as in (7.46).

(7.46) Wie is (er) aan de lijn?
who is (it+R) on the line
'Who is calling?'

The obligatory omission in (7.45), for its part, is due to a constraint which blocks adjacent instances of *er/d'r*. It can be compared to the avoidance of two adjacent instances of the complementizer *of* 'whether' in embedded interrogative clauses like (7.47a), taken from Ackema (2001). He points out that sequences of identical morphemes or heads sometimes result in ill-formed constructions. If the complementizer *dat* 'that' is used, the construction is grammatical (7.47b).[11]

(7.47) a. * Vroeg je nou of die plaats bezet is of of hij vrij is.
asked you now wheter that seat occupied is or whether he free is

b. Vroeg je nou of die plaats bezet is of dat hij vrij is.
asked you now wheter that seat occupied is or that he free is
'Did you ask whether that seat is occupied or whether it is free?'

Similar remarks apply to clauses which contain the locative *er* 'there', such as (7.48).

(7.48) We gaan er de ontsnapte papegaai met een groot net vangen.
we go there the escaped parrot with a large net catch
'We are going to catch the escaped parrot there with a large net.'

If the locative *er* is followed by the homophonous non-nominative R-pronoun, as in (7.49), the latter is omitted.

(7.49) We gaan er (*er) de ontsnapte papegaai [__ mee] vangen.
we go there (*it+R) the escaped parrot with catch
'We are going to catch the escaped parrot there with it.'

In sum, there is no need for amalgamation of syntactic functions, since the relevant data can be modelled in terms of the omissibility of the (nominal) complements of adpositions.

[10] A similar assumption is made in the transformational treatment of Bennis (1986).

[11] Ackema (2001) proposes a substitution analysis of the second *of* by *dat* 'that' in order to get the grammatical variant.

7.3 Conclusion

In this chapter it is argued that the CRP not only accounts for complement raising out of verbal complements: It also deals with complement raising out of adjectival and adpositional complements, as well as with complement raising out of adjuncts and subjects.

The existing HPSG treatments of adposition stranding in Dutch provide a uniform extraction analysis, employing non-local devices as SLASH and BIND, see Rentier (1993) and Bouma (2000). This is adopted here as well for the cases in which the extracted pronouns end up in the Vorfeld, but not for the cases in which they end up in the left part of the Mittelfeld, as the latter concern a medium-distance (bounded) dependency.

In order to avoid overgeneration, the constraint that blocks complement raising out of V-initial VPs and (C-initial) CPs is extended to one that blocks complement raising out of P-initial PPs. Furthermore, a constraint is added to block complement extraction out of P-initial PPs. Having spelled out the treatment, Bouma's objections against the use of argument inheritance for the analysis of adposition stranding were discussed, and it was demonstrated that none of them sticks.

The resulting treatment is not only economical, it also accounts for the fact that languages which abide by the Empty COMPS Constraint, such as English, lack not only the kind of scrambling that we find in Dutch and German, but also the kind of adposition stranding that results from complement raising, as opposed to the kind of adposition stranding that results from complement extraction.

Conclusion

This dissertation addresses a quintessentially Dutch phenomenon, i.e. complement raising and the formation of verb clusters. In addition, it aims to illustrate how a treebank-based study can shed new light on the theoretical analysis of verb clusters and related phenomena, such as word order variation and cluster creeping.

Literature study

The literature study shows that, even though the research on verb clusters is abundant, authors do not entirely agree what exactly constitutes a cluster. Moreover, the set of verbs that (obligatorily or optionally) act as clustering verbs is not uniformly defined.

Besides the descriptive literature on verb clusters, several theoretical accounts are discussed. Transformational approaches start from the assumption that verbs are aligned with their arguments in the underlying structure, to which several transformations are applied in order to derive verb clusters. Instead of modelling verb clusters in terms of overt movement, HPSG deals with verb clusters in terms of the inheritance of unsaturated valence requirements. Pioneering work on the analysis of (German) verb clusters is the argument inheritance treatment proposed in Hinrichs & Nakazawa (1994). It extends the subject raising analysis to a more general mechanism, raising subjects and complements in a similar fashion. This generalized raising approach is adopted and extended by several authors working on German syntax, and it is also applied to analyse Dutch verb clusters.

A treebank-supported investigation

For the corpus study syntactically annotated corpora or treebanks are used, since they allow for the empirical investigation of Dutch syntax beyond the lexical level. The aim of the treebank investigation is to empirically verify the descriptive literature and

the theoretical analyses. The treebank study is split up into several case studies, each focusing on a particular topic related to complement raising and cluster formation.

First, treebank data are collected in order to investigate the general occurrence of verb clusters, as well as the word order variation observed in the data. Special attention goes out to clusters containing *te*-infinitives, as they are often neglected in studies on verb clusters. By investigating all constructions in which a verb selects a *te*-infinitive as a complement in the treebanks, it turns out that the *te*-infinitives that appear to the left of their selector are never instances of clustering verbs, but rather instances of predicative infinitives or parts of idiomatic expressions.

Second, the set of clustering verbs is extracted from the treebanks and compared to similar lists encountered in the literature. In addition, the relation between clustering verbs and the occurrence of the IPP effect is discussed. The data confirm that IPP is an important diagnostic for the identification of clustering verbs, although there is a set of clustering verbs that cannot occur as IPP. While the extraction of verb clusters in general illustrates that verb cluster formation is a frequently occurring phenomenon, the treebank investigation of clustering verbs suffers from data sparseness. The treebanks do not contain instances of all verbs mentioned in the literature on verb clusters. In order to accommodate for this, evidence for their status as a clustering verb is found in data from the Web.

Third, the treebanks were used to investigate cluster creeping, i.e. the interruption of a verb cluster by non-verbal elements. As expected, the phenomenon is more common in spoken Dutch than in written Dutch. The treebank investigation furthermore results in a classification of cluster creepers according to their lexical category and to their syntactic function in the sentence.

Differentiating complement raising from subject raising

An evaluation of the literature study reveals that the argument inheritance analysis is problematic in a number of constructions. For instance, it does not account for complement raising without subject raising. In addition, it poses problems in its interaction with the binding principles and the application of the passive lexical rule.

It is motivated that Dutch verb clusters can be analysed more adequately by treating subject raising and complement raising as separate mechanisms. While the subject requirements of the selected verbs are shared with the subject requirement of their selector, the complement requirements are not. Instead, the unsaturated complement

list of the selected verb is directly propagated to the mother node.

There is both intra- and interlingual evidence for such an analysis. In Dutch, subject raising does not necessarily occur together with complement raising (e.g. in the case of subject control verbs). From a cross-linguistic perspective, there turn out to be languages that allow subject raising, but not complement raising (e.g. English). An important consequence of the new analysis is that it not only deals with complement raising out of verbal complements: It also extends to complement raising out of non-verbal complements, such as adjectival and prepositional phrases. The latter leads to adposition stranding. This phenomenon is spelled out in further detail, showing that raising is possible out of P-final PPs, but not out of P-initial PPs. Furthermore, it was argued that cases of complement raising should be differentiated from complement extraction.

The analysis of complement raising and cluster formation is supported by treebank examples wherever possible. Especially the analysis of word order variation and cluster creeping heavily relies on observations from the treebank data. Even though the extraction of the relevant constructions is not a trivial task, it illustrates that corpus data provide a valuable means for the description and analysis of linguistic phenomena.

Future work

While the new analysis of complement raising deals with typical verb clusters as well as constructions that are ambiguous between clustering and non-clustering, there are still a number of constructions that were not addressed in this dissertation. For instance, future work should investigate whether the complement raising analysis adequately accounts for constructions with coordinated verb clusters.

In addition, the conditions and constraints on complement raising out of non-verbal phrases should be investigated in more detail. The main purpose of this dissertation was to provide an analysis of raising out of verb phrases. It was shown how the complement raising principle extends to other phenomena, such as raising out of APs, NPs, and PPs. The latter was investigated in more detail, but could also profit from an extensive treebank investigation in order to account for instances in which raising out of PPs can, must or cannot occur.

Furthermore, it would be interesting to investigate to what extent the complement raising principle is applicable to other languages showing complement raising and

cluster formation, such as German and Afrikaans.

Some of the phenomena investigated in the treebank study could also benefit from additional corpus research. For instance, while it was shown that the treebanks are sufficiently large in size in order to describe different types of verb clusters in general, the treebank investigation suffered from data sparseness at several points. Especially with respect to the set of clustering verbs and IPP verbs, not all verbs encountered in the literature study could be retrieved in the data. Using a larger treebank could solve this problem. In addition, a corpus investigation on a larger corpus of the *te*-infinitives would be interesting to verify the findings discussed in this thesis, as those verbs are often neglected in (corpus) studies on verb clusters. For instance, in the case of optionally clustering verbs it could be interesting to investigate which (context) phenomena influence the choice between their clustering and non-clustering use. For this, a much larger treebank is necessary, e.g. SoNaR.

A final topic for future work is an artificial intelligence problem, rather than a linguistic subject. The initial queries used for the treebank study were built by means of the query tool GrETEL, which allows treebank querying by example. Although this tool facilitates the exploitation of treebanks, many queries had to be constructed to extract all the relevant constructions. In order to compensate for this query overload, it would be interesting to investigate how the search engine can generalize over several related patterns, e.g. by generalizing over the queries for lexical and non-lexical constructions.

Appendix A

Abbreviations

This list contains the abbreviations that are used in glosses and typed feature structures. For an overview of the abbreviations and tags that are used in the treebanks, see appendix B.

ACC	accusative case
ADVP	adverb phrase
AP	adjective phrase
ARG-ST	argument structure
ARP	Argument Realization Principle
AVM	attribute-value matrix
CAT	category
CG	Categorial Grammar
CGN	Corpus Gesproken Nederlands (Spoken Dutch Corpus)
COMPS	complements
CP	complementizer phrase
CRP	Complement Raising Principle
DAT	dative case
HD-DTR	head daughter
DIM	diminutive form
ECC	Empty Comps Constraint
FIN	finite verb form
GPSG	Generalized Phrase Structure Grammar
HFP	Head Feature Principle
HPSG	Head-driven Phrase Structure Grammar

INF	bare infinitive
IPP	infinitive showing the IPP effect
LASSY	Large Scale Syntactic Annotation of written Dutch (Written Dutch Corpus)
LOC	local
NOM	nominative case
NONHD-DTR	non-head daughter
NP	noun phrase
PHON	phonology
PP	prepositional phrase
POS	part-of-speech
PRP	present participle
PSP	past participle
+R	R-pronoun
RED	reduced form
SUBCAT	subcategorization
SUBJ	subject
SYNSEM	syntax-semantics
TE INF	*te*-infinitive
TFS	typed feature structure
VC	verbal complement
VFORM	verb form
VP	verb phrase

APPENDIX **B**

Treebank annotations

B.1 CGN treebank

This annex provides an overview of the linguistic information in the CGN Treebank (version 2.0.1), following the guidelines in Hoekstra et al. (2003).

B.1.1 Syntactic annotations

The syntactic annotations in the CGN Treebank include the dependency relations (indicated by the label `rel`) and the syntactic categories (indicated by the label `cat`). The Tables B.1 and B.2, respectively list, describe, and provide counts for the possible values for the `rel` and `cat` labels in the CGN Treebank.

TAG	DESCRIPTION	FREQUENCY
–	daughter of a 'top' node	333692
APP	apposition	3414
BODY	sister of a complementizer or a WH-phrase	40629
CMP	complementizer	27917
CNJ	conjunct	38366
CRD	coordinator	19860
DET	determiner	87308
DLINK	discourse link	23286
DP	discourse part	20740
HD	syntactic head	346884
HDF	final part of a circumposition	602
LD	location or direction complement	14782

continued on next page

TAG	DESCRIPTION	FREQUENCY
LP	part of a list	5243
ME	measure complement	383
MOD	modifier	190966
MWP	part of a multi word unit	40175
NUCL	nuclear clause	53603
OBCOMP	comparsion complement	1261
OBJ1	direct object	116478
OBJ2	indirect or secondary object	2098
PART	partitive	5
PC	prepositional complement	11279
POBJ1	preliminary direct object	633
PREDC	predicative complement	28396
PREDM	predicative modifier	3244
PRT	adverbial particle	10179
RHD	head of antecedentless relative clause	7579
SAT	satellite	5692
SE	obligatory reflexive object	1271
SU	subject	114206
SUP	preliminary subject	4537
SVP	separable verbal particle	7397
TAG	appendix, parenthesis	44155
TOP	root node of the dependency structure	129923
VC	verbal complement	39233
WHD	head of a WH-question	5147

Table B.1: Dependency relations (`rel`) in the CGN Treebank

TAG	DESCRIPTION	FREQUENCY
–	– (annotation error)	1
advp	adverbial phrase	10475
ahi	*aan het infinitive*	396
ap	adjectival phrase	12480
conj	coordinate phrase with conjunction	18059
cp	complementizer phrase	17761
detp	determiner phrase	2715
du	discourse unit	64442
inf	bare infinitival phrase	26293
list	coordinate phrase without conjunction	2111
mwu	multi word unit	17769
np	noun phrase	102734
oti	*om te infinitive*	2741
pp	prepositional phrase	66154
ppart	past participial phrase	12501
ppres	present participial phrase	171
rel	relative clause	5953
smain	main clause (SVO)	77313
ssub	subordinate clause (SOV)	25466
sv1	verb initial clause (VSO)	15910
svan	*van clause*	1397
ti	*te infinitive*	5623
top	root of a dependency structure	129923
whq	main clause wh-question	3532
whrel	free relative	1642
whsub	subordinate clause wh-question	1606

Table B.2: Syntactic categories (`cat`) in the CGN Treebank (non-lexical nodes)

B.1.2 Lexical annotations

The lexical annotations in the CGN Treebank include CGN/D-Coi POS tags, following Van Eynde (2004). The CGN tag set is designed for spoken Dutch, whereas the D-Coi tag set is an adaptation of the CGN tag set, focusing on written Dutch.

The full CGN/D-Coi tags are indicated by the label `postag`, whereas the shortened version is labeled `pt`. Table B.3 lists, describes, and provides counts for the values of the `pt` label. Table B.4 lists and provides counts for the values of the `postag` label.

Descriptions and examples of those tags can be found in Van Eynde (2004). Note that the counts of each tag in Table B.3 are the sum of the subtypes listed in Table B.4.

TAG	DESCRIPTION	FREQUENCY
adj	adjective	63764
bw	adverb	120277
let	punctuation	129923
lid	article	59179
n	noun	126199
spec	special token	37188
tsw	interjection	88493
tw	numeral	14610
vg	conjunction	64469
vnw	pronoun	177101
vz	preposition	88153
ww	verb	170538

Table B.3: Short CGN/D-Coi POS tags (`pt`) in the CGN Treebank (lexical nodes)

TAG	FREQUENCY
ADJ(dial)	86
ADJ(nom,basis,met-e,mv-n)	295
ADJ(nom,basis,met-e,zonder-n,bijz)	2
ADJ(nom,basis,met-e,zonder-n,stan)	615
ADJ(nom,basis,zonder,mv-n)	6
ADJ(nom,basis,zonder,zonder-n)	544
ADJ(nom,comp,met-e,mv-n)	56
ADJ(nom,comp,met-e,zonder-n,stan)	11
ADJ(nom,comp,zonder,zonder-n)	2
ADJ(nom,sup,met-e,mv-n)	9
ADJ(nom,sup,met-e,zonder-n,stan)	270
ADJ(nom,sup,zonder,zonder-n)	67
ADJ(postnom,basis,met-s)	451
ADJ(postnom,basis,zonder)	90
ADJ(postnom,comp,met-s)	7
ADJ(postnom,comp,zonder)	5
ADJ(prenom,basis,met-e,bijz)	10
ADJ(prenom,basis,met-e,stan)	12991
ADJ(prenom,basis,zonder)	5001

continued on next page

TAG	FREQUENCY
ADJ(prenom,comp,met-e,stan)	282
ADJ(prenom,comp,zonder)	99
ADJ(prenom,sup,met-e,stan)	838
ADJ(prenom,sup,zonder)	5
ADJ(vrij,basis,zonder)	38907
ADJ(vrij,comp,zonder)	2640
ADJ(vrij,dim,zonder)	80
ADJ(vrij,sup,zonder)	395
BW()	120269
BW(dial)	8
LET()	129923
LID(bep,dat,evmo)	23
LID(bep,dial)	359
LID(bep,gen,evmo)	470
LID(bep,gen,rest3)	40
LID(bep,stan,evon)	10556
LID(bep,stan,rest)	27490
LID(onbep,dial)	602
LID(onbep,stan,agr)	19639
N(eigen,dial)	1
N(eigen,ev,basis,gen)	113
N(eigen,ev,basis,genus,stan)	166
N(eigen,ev,basis,onz,stan)	6255
N(eigen,ev,basis,zijd,stan)	11258
N(eigen,ev,dim,onz,stan)	30
N(eigen,mv,basis)	712
N(eigen,mv,dim)	1
N(soort,dial)	30
N(soort,ev,basis,dat)	327
N(soort,ev,basis,gen)	461
N(soort,ev,basis,genus,stan)	717
N(soort,ev,basis,onz,stan)	25637
N(soort,ev,basis,zijd,stan)	51999
N(soort,ev,dim,onz,stan)	4522
N(soort,mv,basis)	22584
N(soort,mv,dim)	1386
SPEC(afgebr)	9235
SPEC(deeleigen)	10101
SPEC(meta)	360
SPEC(onverst)	14992

continued on next page

TAG	FREQUENCY
SPEC(vreemd)	2500
TSW()	88491
TSW(dial)	2
TW(hoofd,dial)	44
TW(hoofd,nom,mv-n,basis)	114
TW(hoofd,nom,mv-n,dim)	1
TW(hoofd,nom,zonder-n,basis)	409
TW(hoofd,nom,zonder-n,dim)	131
TW(hoofd,prenom,bijz)	4
TW(hoofd,prenom,stan)	7246
TW(hoofd,vrij)	5152
TW(rang,dial)	13
TW(rang,nom,mv-n)	5
TW(rang,nom,zonder-n)	311
TW(rang,prenom,bijz)	1
TW(rang,prenom,stan)	1179
VG(neven)	44999
VG(onder)	19465
VG(onder,dial)	5
VNW(aanw,adv-pron,obl,vol,3o,getal)	7459
VNW(aanw,adv-pron,stan,red,3,getal)	7042
VNW(aanw,det,dat,nom,met-e,zonder-n)	1
VNW(aanw,det,dat,prenom,met-e,evmo)	1
VNW(aanw,det,dial)	155
VNW(aanw,det,gen,prenom,met-e,rest3)	2
VNW(aanw,det,stan,nom,met-e,mv-n)	32
VNW(aanw,det,stan,nom,met-e,zonder-n)	178
VNW(aanw,det,stan,prenom,met-e,rest)	868
VNW(aanw,det,stan,prenom,zonder,agr)	1271
VNW(aanw,det,stan,prenom,zonder,evon)	2797
VNW(aanw,det,stan,prenom,zonder,rest)	6862
VNW(aanw,det,stan,vrij,zonder)	7
VNW(aanw,pron,dial)	49
VNW(aanw,pron,gen,vol,3m,ev)	5
VNW(aanw,pron,gen,vol,3o,ev)	2
VNW(aanw,pron,stan,nadr,3o,ev)	3
VNW(aanw,pron,stan,vol,3,getal)	5538
VNW(aanw,pron,stan,vol,3o,ev)	21404
VNW(betr,det,stan,nom,met-e,zonder-n)	4
VNW(betr,det,stan,nom,zonder,zonder-n)	25

continued on next page

TAG	FREQUENCY
VNW(betr,pron,dial)	7
VNW(betr,pron,gen,vol,3o,ev)	1
VNW(betr,pron,stan,vol,3,ev)	525
VNW(betr,pron,stan,vol,persoon,getal)	3870
VNW(bez,det,dat,vol,3,ev,prenom,met-e,evf)	1
VNW(bez,det,dial)	190
VNW(bez,det,gen,vol,1,ev,prenom,zonder,evmo)	1
VNW(bez,det,gen,vol,3,ev,prenom,met-e,rest3)	1
VNW(bez,det,gen,vol,3,ev,prenom,zonder,evmo)	2
VNW(bez,det,stan,nadr,2v,mv,prenom,zonder,agr)	60
VNW(bez,det,stan,red,1,ev,prenom,zonder,agr)	846
VNW(bez,det,stan,red,2v,ev,prenom,zonder,agr)	824
VNW(bez,det,stan,red,3,ev,prenom,zonder,agr)	1065
VNW(bez,det,stan,red,3,getal,prenom,zonder,agr)	121
VNW(bez,det,stan,vol,1,ev,nom,met-e,zonder-n)	11
VNW(bez,det,stan,vol,1,ev,prenom,met-e,rest)	6
VNW(bez,det,stan,vol,1,ev,prenom,zonder,agr)	1164
VNW(bez,det,stan,vol,1,mv,nom,met-e,mv-n)	1
VNW(bez,det,stan,vol,1,mv,prenom,met-e,rest)	436
VNW(bez,det,stan,vol,1,mv,prenom,zonder,evon)	289
VNW(bez,det,stan,vol,2,getal,nom,met-e,zonder-n)	1
VNW(bez,det,stan,vol,2,getal,prenom,zonder,agr)	455
VNW(bez,det,stan,vol,2v,ev,nom,met-e,zonder-n)	1
VNW(bez,det,stan,vol,2v,ev,prenom,zonder,agr)	176
VNW(bez,det,stan,vol,3,ev,prenom,zonder,agr)	1192
VNW(bez,det,stan,vol,3,mv,prenom,zonder,agr)	656
VNW(bez,det,stan,vol,3m,ev,nom,met-e,zonder-n)	3
VNW(bez,det,stan,vol,3p,mv,prenom,met-e,rest)	1
VNW(bez,det,stan,vol,3v,ev,nom,met-e,zonder-n)	3
VNW(excl,det,stan,vrij,zonder)	1
VNW(excl,pron,stan,vol,3,getal)	172
VNW(onbep,adv-pron,gen,red,3,getal)	104
VNW(onbep,adv-pron,obl,vol,3o,getal)	486
VNW(onbep,det,dat,prenom,met-e,evmo)	1
VNW(onbep,det,dial)	69
VNW(onbep,det,gen,nom,met-e,mv-n)	2
VNW(onbep,det,gen,prenom,met-e,mv)	3
VNW(onbep,det,stan,nom,met-e,mv-n)	104
VNW(onbep,det,stan,nom,met-e,zonder-n)	34
VNW(onbep,det,stan,nom,zonder,zonder-n)	84
VNW(onbep,det,stan,prenom,met-e,agr)	538

continued on next page

TAG	FREQUENCY
VNW(onbep,det,stan,prenom,met-e,evz)	550
VNW(onbep,det,stan,prenom,met-e,mv)	25
VNW(onbep,det,stan,prenom,met-e,rest)	352
VNW(onbep,det,stan,prenom,zonder,agr)	2006
VNW(onbep,det,stan,prenom,zonder,evon)	708
VNW(onbep,det,stan,vrij,zonder)	451
VNW(onbep,grad,gen,nom,met-e,mv-n,basis)	1
VNW(onbep,grad,stan,nom,met-e,mv-n,basis)	38
VNW(onbep,grad,stan,nom,met-e,mv-n,sup)	20
VNW(onbep,grad,stan,nom,met-e,zonder-n,basis)	4
VNW(onbep,grad,stan,nom,met-e,zonder-n,sup)	22
VNW(onbep,grad,stan,prenom,met-e,agr,basis)	52
VNW(onbep,grad,stan,prenom,met-e,agr,comp)	5
VNW(onbep,grad,stan,prenom,met-e,agr,sup)	103
VNW(onbep,grad,stan,prenom,met-e,mv,basis)	57
VNW(onbep,grad,stan,prenom,zonder,agr,basis)	1010
VNW(onbep,grad,stan,prenom,zonder,agr,comp)	478
VNW(onbep,grad,stan,vrij,zonder,basis)	1407
VNW(onbep,grad,stan,vrij,zonder,comp)	2349
VNW(onbep,grad,stan,vrij,zonder,sup)	99
VNW(onbep,pron,dial)	7
VNW(onbep,pron,gen,vol,3p,ev)	8
VNW(onbep,pron,stan,vol,3o,ev)	5126
VNW(onbep,pron,stan,vol,3p,ev)	1188
VNW(pers,pron,dial)	826
VNW(pers,pron,gen,vol,1,mv)	2
VNW(pers,pron,nomin,nadr,1,ev)	43
VNW(pers,pron,nomin,nadr,2b,getal)	2
VNW(pers,pron,nomin,nadr,2v,ev)	1
VNW(pers,pron,nomin,nadr,3m,ev,masc)	4
VNW(pers,pron,nomin,nadr,3p,mv)	1
VNW(pers,pron,nomin,nadr,3v,ev,fem)	3
VNW(pers,pron,nomin,red,1,ev)	4819
VNW(pers,pron,nomin,red,1,mv)	5680
VNW(pers,pron,nomin,red,2,getal)	1370
VNW(pers,pron,nomin,red,2v,ev)	11820
VNW(pers,pron,nomin,red,3,ev,masc)	1987
VNW(pers,pron,nomin,red,3p,ev,masc)	461
VNW(pers,pron,nomin,vol,1,ev)	20635
VNW(pers,pron,nomin,vol,1,mv)	1828
VNW(pers,pron,nomin,vol,2,getal)	403

continued on next page

TAG	FREQUENCY
VNW(pers,pron,nomin,vol,2b,getal)	1619
VNW(pers,pron,nomin,vol,2v,ev)	1198
VNW(pers,pron,nomin,vol,3,ev,masc)	2865
VNW(pers,pron,nomin,vol,3p,mv)	407
VNW(pers,pron,nomin,vol,3v,ev,fem)	393
VNW(pers,pron,obl,nadr,3m,ev,masc)	3
VNW(pers,pron,obl,nadr,3p,mv)	5
VNW(pers,pron,obl,red,3,ev,masc)	779
VNW(pers,pron,obl,red,3v,getal,fem)	75
VNW(pers,pron,obl,vol,2v,ev)	372
VNW(pers,pron,obl,vol,3,ev,masc)	537
VNW(pers,pron,obl,vol,3,getal,fem)	321
VNW(pers,pron,obl,vol,3p,mv)	216
VNW(pers,pron,stan,nadr,2v,mv)	553
VNW(pers,pron,stan,red,3,ev,fem)	2226
VNW(pers,pron,stan,red,3,ev,onz)	14187
VNW(pers,pron,stan,red,3,mv)	4744
VNW(pr,pron,obl,nadr,1,ev)	83
VNW(pr,pron,obl,nadr,1,mv)	11
VNW(pr,pron,obl,nadr,2,getal)	4
VNW(pr,pron,obl,nadr,2v,getal)	73
VNW(pr,pron,obl,red,1,ev)	973
VNW(pr,pron,obl,red,2v,getal)	525
VNW(pr,pron,obl,vol,1,ev)	1771
VNW(pr,pron,obl,vol,1,mv)	751
VNW(pr,pron,obl,vol,2,getal)	417
VNW(recip,pron,gen,vol,persoon,mv)	5
VNW(recip,pron,obl,vol,persoon,mv)	654
VNW(refl,pron,dial)	6
VNW(refl,pron,obl,nadr,3,getal)	113
VNW(refl,pron,obl,red,3,getal)	1030
VNW(vb,adv-pron,obl,vol,3o,getal)	1608
VNW(vb,det,stan,nom,met-e,zonder-n)	11
VNW(vb,det,stan,prenom,met-e,rest)	267
VNW(vb,det,stan,prenom,zonder,evon)	95
VNW(vb,pron,gen,vol,3m,ev)	6
VNW(vb,pron,stan,vol,3o,ev)	4519
VNW(vb,pron,stan,vol,3p,getal)	489
VNW(vrag,det,dial)	5
VNW(vrag,pron,dial)	5

continued on next page

TAG	FREQUENCY
VNW(vrag,pron,stan,nadr,3o,ev)	17
VZ(fin)	12253
VZ(fin,dial)	1
VZ(init)	75607
VZ(versm)	292
WW(dial)	45
WW(inf,nom,zonder,zonder-n)	1593
WW(inf,prenom,zonder)	1
WW(inf,vrij,zonder)	30147
WW(od,nom,met-e,mv-n)	2
WW(od,nom,met-e,zonder-n)	42
WW(od,prenom,met-e)	678
WW(od,prenom,zonder)	193
WW(od,vrij,zonder)	357
WW(pv,conj,ev)	75
WW(pv,tgw,ev)	54921
WW(pv,tgw,met-t)	23355
WW(pv,tgw,mv)	17317
WW(pv,verl,ev)	20522
WW(pv,verl,met-t)	109
WW(pv,verl,mv)	4419
WW(vd,nom,met-e,mv-n)	23
WW(vd,nom,met-e,zonder-n)	30
WW(vd,prenom,met-e)	363
WW(vd,prenom,zonder)	462
WW(vd,vrij,zonder)	15884

Table B.4: CGN/D-Coi POS tags (`postag`) in the CGN Treebank (lexical nodes)

B.1.3 Data format

The annotations are encoded in the XML structure of the treebank. The original data format of the CGN Treebank is TIGER-XML.[1] In order to make the treebank queryable with the same tools as the LASSY treebank, the CGN Treebank was converted by Gertjan van Noord into the Alpino-XML data format.[2] In addition to the conversion,

[1] `http://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/TIGERSearch/doc/html/TigerXML.html`

[2] `http://www.let.rug.nl/vannoord/Lassy/alpino_ds.dtd`

lemmas were added as well.

The TIGER-XML source code of the tree in Figure B.1 is given in Figure B.2, and the Alpino-XML conversion is presented in B.3.

In the TIGER-XML format, lexical information (included in the terminal nodes) is separated from nonlexical information (non-terminals).

In the Alpino-XML format, every node in the tree is represented by a `node` element in the XML structure, which includes both lexical and non-lexical information. The annotations presented in the tables above are represented as XML attribute-value pairs in those nodes. The way the nodes are hierarchically organized in Alpino-XML corresponds to the nodes in the graphical representation in Figure B.1, whereas there is no such correspondence in the TIGER-XML format. Note that the graphical tree only shows some of the features encoded in the XML data.

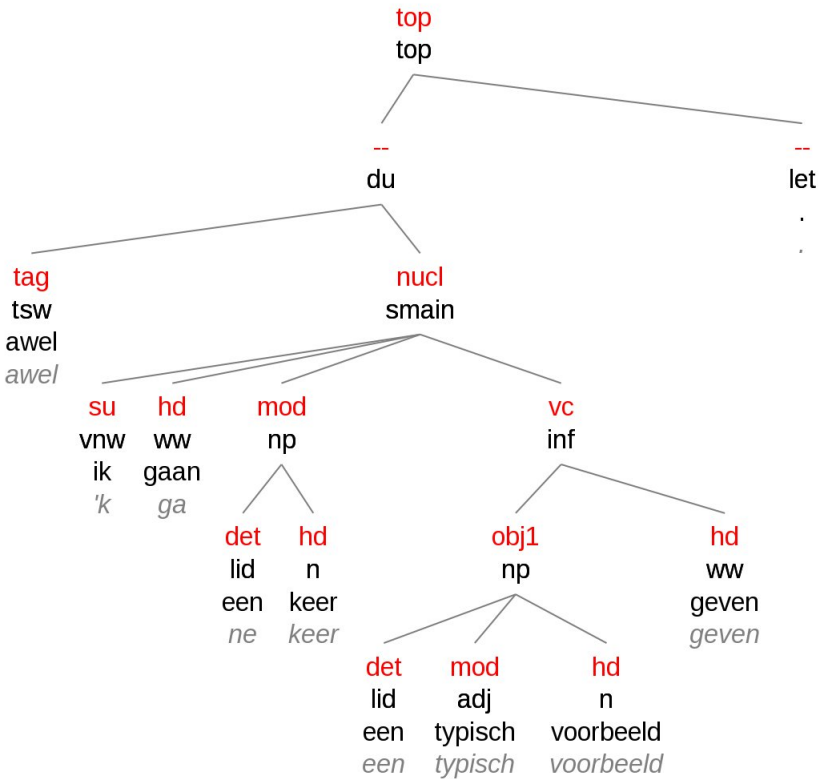


Figure B.1: Tree representation of a CGN sentence [fva400392__6]

```
<s id="fv400392.6">
<graph root="fv400392.6.504">
<terminals>
 <t id="fv400392.6.1" word="awel" pos="TSW" morph="T001" />
 <t id="fv400392.6.2" word="'k" pos="VNW1" morph="T501c" />
 <t id="fv400392.6.3" word="ga" pos="WW1" morph="T301" />
 <t id="fv400392.6.4" word="ne" pos="LID" morph="R602" />
 <t id="fv400392.6.5" word="keer" pos="N1" morph="U117" />
 <t id="fv400392.6.6" word="een" pos="LID" morph="U608" />
 <t id="fv400392.6.7" word="typisch" pos="ADJ1" morph="T201" />
 <t id="fv400392.6.8" word="voorbeeld" pos="N1" morph="T102" />
 <t id="fv400392.6.9" word="geven" pos="WW4" morph="T314" />
 <t id="fv400392.6.10" word="." pos="LET" morph="T007" />
</terminals>
<nonterminals>
 <nt id="fv400392.6.500" cat="NP">
  <edge label="DET" idref="fv400392.6.4" />
  <edge label="HD" idref="fv400392.6.5" />
 </nt>
 <nt id="fv400392.6.501" cat="NP">
  <edge label="DET" idref="fv400392.6.6" />
  <edge label="MOD" idref="fv400392.6.7" />
  <edge label="HD" idref="fv400392.6.8" />
 </nt>
 <nt id="fv400392.6.502" cat="INF">
  <edge label="HD" idref="fv400392.6.9" />
  <edge label="OBJ1" idref="fv400392.6.501" />
 </nt>
 <nt id="fv400392.6.503" cat="SMAIN">
  <edge label="SU" idref="fv400392.6.2" />
  <edge label="HD" idref="fv400392.6.3" />
  <edge label="MOD" idref="fv400392.6.500" />
  <edge label="VC" idref="fv400392.6.502" />
 </nt>
 <nt id="fv400392.6.504" cat="DU">
  <edge label="TAG" idref="fv400392.6.1" />
  <edge label="NUCL" idref="fv400392.6.503" />
 </nt>
</nonterminals>
</graph>
</s>
```

Figure B.2: TIGER-XML structure of a CGN sentence [fva400392__6]

```
<alpino_ds id="fva400392__6" version="1.3" >
  <node begin="0" cat="top" end="10" id="0" rel="top">
    <node begin="0" cat="du" end="9" id="1" rel="--">
      <node begin="0" end="1" id="2" lcat="TSW()" pos="T001" postag="TSW()" pt="tsw"
      rel="tag" root="awel" word="awel" lemma="awel"/>
      <node begin="1" cat="smain" end="9" id="3" rel="nucl">
        <node begin="1" end="2" getal="ev" id="4" lcat="VNW(pers,pron,nomin,red,1,ev)"
        naamval="nomin" pdtype="pron" persoon="1" pos="T501c"
        postag="VNW(pers,pron,nomin,red,1,ev)" pt="vnw" rel="su" root="'k" status="red"
        vwtype="pers" word="'k" lemma="ik"/>
        <node begin="2" end="3" id="5" lcat="WW(pv,tgw,ev)" pos="T301"
        postag="WW(pv,tgw,ev)" pt="ww" pvagr="ev" pvtijd="tgw" rel="hd" root="ga"
        word="ga" wvorm="pv" lemma="gaan"/>
        <node begin="3" cat="np" end="5" id="6" rel="mod">
          <node begin="3" dial="dial" end="4" id="7" lcat="LID(onbep,dial)"
          lwtype="onbep" pos="R602" postag="LID(onbep,dial)" pt="lid" rel="det"
          root="ne" word="ne" lemma="een"/>
          <node begin="4" end="5" genus="genus" getal="ev" graad="basis" id="8"
          lcat="N(soort,ev,basis,genus,stan)" naamval="stan" ntype="soort" pos="U117"
          postag="N(soort,ev,basis,genus,stan)" pt="n" rel="hd" root="keer" word="keer"
          lemma="keer"/>
        </node>
        <node begin="5" cat="inf" end="9" id="9" rel="vc">
          <node begin="5" cat="np" end="8" id="10" rel="obj1">
            <node begin="5" end="6" id="11" lcat="LID(onbep,stan,agr)" lwtype="onbep"
            naamval="stan" npagr="agr" pos="U608" postag="LID(onbep,stan,agr)" pt="lid"
            rel="det" root="een" word="een" lemma="een"/>
            <node begin="6" buiging="zonder" end="7" graad="basis" id="12"
            lcat="ADJ(prenom,basis,zonder)" pos="T201" positie="prenom"
            postag="ADJ(prenom,basis,zonder)" pt="adj" rel="mod" root="typisch"
            word="typisch" lemma="typisch"/>
            <node begin="7" end="8" genus="onz" getal="ev" graad="basis" id="13"
            lcat="N(soort,ev,basis,onz,stan)" naamval="stan" ntype="soort" pos="T102"
            postag="N(soort,ev,basis,onz,stan)" pt="n" rel="hd" root="voorbeeld"
            word="voorbeeld" lemma="voorbeeld"/>
          </node>
          <node begin="8" buiging="zonder" end="9" id="14" lcat="WW(inf,vrij,zonder)"
          pos="T314" positie="vrij" postag="WW(inf,vrij,zonder)" pt="ww" rel="hd"
          root="geven" word="geven" wvorm="inf" lemma="geven"/>
        </node>
      </node>
    </node>
    <node begin="9" end="10" id="15" lcat="LET()" pos="T007" postag="LET()" pt="let"
    rel="--" root="." word="." lemma="."/>
  </node>
  <sentence>awel 'k ga ne keer een typisch voorbeeld geven .</sentence>
</alpino_ds>
```

Figure B.3: Alpino-XML structure of a CGN sentence [fva400392__6]

B.2 LASSY Small

This annex provides an overview of the linguistic information in LASSY Small (version 1.1), following the LASSY annotation manual (van Noord et al. 2011).

B.2.1 Syntactic annotations

The syntactic annotations in LASSY Small comprise the dependency relations (indicated by the label `rel`) and the syntactic categories (indicated by the label `cat`). The

Tables B.5 and B.6 respectively list, describe, and provide counts for the values of the `rel` and `cat` labels in LASSY Small.

TAG	DESCRIPTION	FREQUENCY
–	daughter of a 'top' node	187345
app	apposition	11955
body	sister of a complementizer or a WH-phrase	43242
cmp	complementizer	30785
cnj	conjunct	66638
crd	coordinator	29063
det	determiner	149466
dlink	discourse link	2989
dp	discourse part	12405
hd	syntactic head	460145
hdf	final part of a circumposition	289
ld	location or direction complement	7884
me	measure complement	1456
mod	modifier	253909
mwp	part of a multi word unit	78892
nucl	nuclear clause	7999
obcomp	comparsion complement	1719
obj1	direct object	166357
obj2	indirect or secondary object	2479
pc	prepositional complement	16324
pobj1	preliminary direct object	553
predc	predicative complement	19452
predm	predicative modifier	2191
rhd	head of antecedentless relative clause	10815
sat	satellite	2872
se	obligatory reflexive object	1855
su	subject	106625
sup	preliminary subject	948
svp	separable verbal particle	10327
tag	appendix, parenthesis	2154
top	root node of the dependency structure	65200
vc	verbal complement	43119
whd	head of a WH-question	1645

Table B.5: Dependency relations (`rel`) in LASSY Small

TAG	DESCRIPTION	FREQUENCY
advp	adverbial phrase	3180
ahi	*aan het infinitive*	63
ap	adjectival phrase	11024
conj	coordinate phrase with conjunction	29781
cp	complementizer phrase	16938
detp	determiner phrase	1198
du	discourse unit	13696
inf	bare infinitival phrase	24179
mwu	multi word unit	30575
np	noun phrase	190626
oti	*om te infinitive*	4314
pp	prepositional phrase	120889
ppart	past participial phrase	21621
ppres	present participial phrase	626
rel	relative clause	9700
smain	main clause (SVO)	58606
ssub	subordinate clause (SOV)	24211
sv1	verb initial clause (VSO)	3989
svan	*van clause*	16
ti	*te infinitive*	9456
top	root of a dependency structure	65200
whq	main clause wh-question	619
whrel	free relative	1115
whsub	subordinate clause wh-question	1026

Table B.6: Syntactic categories (`cat`) in LASSY Small (non-lexical nodes)

B.2.2 Lexical annotations

The lexical annotations in LASSY Small include the Alpino POS tags (indicated by the label `pos`),[3] and the CGN/D-Coi POS tags, following Van Eynde (2004). The full CGN/D-Coi tags are indicated by the label `postag`, whereas the shortened version is labeled `pt`.

The Tables B.7 and B.8 list, describe, and provide counts for the values of the `pos` and `pt` labels. Table B.9 lists and provides counts for the values of the `postag` label. Descriptions and examples of those tags can be found in Van Eynde (2005). The counts of each tag in Table B.8 are the sum of the subtypes listed in Table B.9.

[3] `http://www.let.rug.nl/vannoord/alp/Alpino/adt.html#postags`

TAG	DESCRIPTION	FREQUENCY
–	–	252
adj	adjective	93227
adv	adverb	50029
comp	complementizer	32886
comparative	comparative complement	1585
det	determiner	144637
fixed	(part of) a fixed expression	3869
name	proper noun	83717
noun	common noun	214262
num	number	18071
part	particle	8043
pp	pronominal adverb	5757
prefix	first conjunct	694
prep	preposition	125634
pron	pronoun	28230
punct	punctuation mark	121122
tag	tag	1259
verb	verb	133491
vg	coordinating conjunction	29412

Table B.7: Alpino POS tags (`pos`) in LASSY Small (lexical nodes)

TAG	DESCRIPTION	FREQUENCY
adj	adjective	80515
bw	adverb	49757
let	punctuation	122360
lid	article	117657
n	noun	246755
spec	special token	39006
tsw	interjection	290
tw	numeral	26454
vg	conjunction	45657
vnw	pronoun	71656
vz	preposition	151123
ww	verb	144947

Table B.8: Short CGN/D-Coi POS tags (`pt`) in LASSY Small (lexical nodes)

TAG	FREQUENCY
ADJ(nom,basis,met-e,mv-n)	869
ADJ(nom,basis,met-e,zonder-n,bijz)	25
ADJ(nom,basis,met-e,zonder-n,stan)	636
ADJ(nom,basis,zonder,mv-n)	11
ADJ(nom,basis,zonder,zonder-n)	514
ADJ(nom,comp,met-e,mv-n)	258
ADJ(nom,comp,met-e,zonder-n,stan)	17
ADJ(nom,sup,met-e,mv-n)	17
ADJ(nom,sup,met-e,zonder-n,stan)	264
ADJ(nom,sup,zonder,zonder-n)	67
ADJ(postnom,basis,met-s)	77
ADJ(postnom,basis,zonder)	55
ADJ(postnom,comp,met-s)	1
ADJ(postnom,comp,zonder)	1
ADJ(prenom,basis,met-e,bijz)	5
ADJ(prenom,basis,met-e,stan)	40171
ADJ(prenom,basis,zonder)	9623
ADJ(prenom,comp,met-e,stan)	1013
ADJ(prenom,comp,zonder)	251
ADJ(prenom,sup,met-e,stan)	1734
ADJ(vrij,basis,zonder)	22165
ADJ(vrij,comp,zonder)	2556
ADJ(vrij,dim,zonder)	19
ADJ(vrij,sup,zonder)	166
BW()	49757
LET()	122360
LID(bep,dat,evf)	1
LID(bep,dat,evmo)	18
LID(bep,gen,evmo)	105
LID(bep,gen,rest3)	172
LID(bep,stan,evon)	25484
LID(bep,stan,rest)	68854
LID(onbep,stan,agr)	23023
N(eigen,ev,basis,gen)	549
N(eigen,ev,basis,genus,stan)	527
N(eigen,ev,basis,onz,stan)	18985
N(eigen,ev,basis,zijd,stan)	18276
N(eigen,ev,dim,onz,stan)	11
N(eigen,mv,basis)	1878
N(soort,ev,basis,dat)	596

continued on next page

TAG	FREQUENCY
N(soort,ev,basis,gen)	122
N(soort,ev,basis,genus,stan)	662
N(soort,ev,basis,onz,stan)	47461
N(soort,ev,basis,zijd,stan)	100103
N(soort,ev,dim,onz,stan)	1005
N(soort,mv,basis)	55983
N(soort,mv,dim)	597
SPEC(afgebr)	851
SPEC(afk)	945
SPEC(deeleigen)	27511
SPEC(enof)	43
SPEC(meta)	18
SPEC(symb)	4234
SPEC(vreemd)	5404
TSW()	290
TW(hoofd,nom,mv-n,basis)	309
TW(hoofd,nom,zonder-n,basis)	697
TW(hoofd,nom,zonder-n,dim)	22
TW(hoofd,prenom,stan)	9783
TW(hoofd,vrij)	12408
TW(rang,nom,mv-n)	24
TW(rang,nom,zonder-n)	422
TW(rang,prenom,bijz)	2
TW(rang,prenom,stan)	2787
VG(neven)	31525
VG(onder)	14132
VNW(aanw,adv-pron,obl,vol,3o,getal)	1355
VNW(aanw,adv-pron,stan,red,3,getal)	4573
VNW(aanw,det,dat,nom,met-e,zonder-n)	1
VNW(aanw,det,dat,prenom,met-e,evmo)	8
VNW(aanw,det,gen,prenom,met-e,rest3)	8
VNW(aanw,det,stan,nom,met-e,mv-n)	25
VNW(aanw,det,stan,nom,met-e,zonder-n)	354
VNW(aanw,det,stan,prenom,met-e,rest)	2688
VNW(aanw,det,stan,prenom,zonder,agr)	324
VNW(aanw,det,stan,prenom,zonder,evon)	1872
VNW(aanw,det,stan,prenom,zonder,rest)	1467
VNW(aanw,det,stan,vrij,zonder)	1
VNW(aanw,pron,gen,vol,3m,ev)	40
VNW(aanw,pron,gen,vol,3o,ev)	3

continued on next page

TAG	FREQUENCY
VNW(aanw,pron,stan,vol,3,getal)	712
VNW(aanw,pron,stan,vol,3o,ev)	3293
VNW(betr,det,stan,nom,zonder,zonder-n)	26
VNW(betr,pron,stan,vol,3,ev)	1285
VNW(betr,pron,stan,vol,persoon,getal)	5799
VNW(bez,det,dat,vol,3,ev,prenom,met-e,evf)	3
VNW(bez,det,gen,vol,1,ev,prenom,zonder,evmo)	2
VNW(bez,det,gen,vol,1,mv,prenom,met-e,rest3)	1
VNW(bez,det,stan,nadr,2v,mv,prenom,zonder,agr)	1
VNW(bez,det,stan,red,1,ev,prenom,zonder,agr)	3
VNW(bez,det,stan,red,2v,ev,prenom,zonder,agr)	198
VNW(bez,det,stan,red,3,ev,prenom,zonder,agr)	196
VNW(bez,det,stan,vol,1,ev,nom,met-e,zonder-n)	2
VNW(bez,det,stan,vol,1,ev,prenom,zonder,agr)	310
VNW(bez,det,stan,vol,1,mv,nom,met-e,zonder-n)	7
VNW(bez,det,stan,vol,1,mv,prenom,met-e,rest)	506
VNW(bez,det,stan,vol,1,mv,prenom,zonder,evon)	192
VNW(bez,det,stan,vol,2,getal,nom,met-e,zonder-n)	1
VNW(bez,det,stan,vol,2,getal,prenom,zonder,agr)	760
VNW(bez,det,stan,vol,2v,ev,prenom,zonder,agr)	16
VNW(bez,det,stan,vol,3,ev,prenom,zonder,agr)	4453
VNW(bez,det,stan,vol,3,mv,prenom,zonder,agr)	2094
VNW(bez,det,stan,vol,3m,ev,nom,met-e,mv-n)	3
VNW(bez,det,stan,vol,3m,ev,nom,met-e,zonder-n)	3
VNW(bez,det,stan,vol,3m,ev,prenom,met-e,rest)	12
VNW(bez,det,stan,vol,3p,mv,nom,met-e,mv-n)	1
VNW(bez,det,stan,vol,3p,mv,nom,met-e,zonder-n)	1
VNW(bez,det,stan,vol,3p,mv,prenom,met-e,rest)	2
VNW(bez,det,stan,vol,3v,ev,prenom,met-e,rest)	12
VNW(excl,pron,stan,vol,3,getal)	2
VNW(onbep,adv-pron,gen,red,3,getal)	40
VNW(onbep,adv-pron,obl,vol,3o,getal)	126
VNW(onbep,det,dat,prenom,met-e,evmo)	3
VNW(onbep,det,gen,nom,met-e,mv-n)	1
VNW(onbep,det,gen,prenom,met-e,mv)	12
VNW(onbep,det,stan,nom,met-e,mv-n)	119
VNW(onbep,det,stan,nom,met-e,zonder-n)	77
VNW(onbep,det,stan,nom,zonder,zonder-n)	2
VNW(onbep,det,stan,prenom,met-e,agr)	1088
VNW(onbep,det,stan,prenom,met-e,evz)	437
VNW(onbep,det,stan,prenom,met-e,mv)	97

continued on next page

TAG	FREQUENCY
VNW(onbep,det,stan,prenom,met-e,rest)	933
VNW(onbep,det,stan,prenom,zonder,agr)	1713
VNW(onbep,det,stan,prenom,zonder,evon)	425
VNW(onbep,det,stan,vrij,zonder)	251
VNW(onbep,grad,gen,nom,met-e,mv-n,basis)	3
VNW(onbep,grad,stan,nom,met-e,mv-n,basis)	121
VNW(onbep,grad,stan,nom,met-e,mv-n,sup)	19
VNW(onbep,grad,stan,nom,met-e,zonder-n,basis)	50
VNW(onbep,grad,stan,nom,met-e,zonder-n,sup)	48
VNW(onbep,grad,stan,nom,zonder,zonder-n,sup)	15
VNW(onbep,grad,stan,prenom,met-e,agr,basis)	307
VNW(onbep,grad,stan,prenom,met-e,agr,comp)	8
VNW(onbep,grad,stan,prenom,met-e,agr,sup)	229
VNW(onbep,grad,stan,prenom,met-e,mv,basis)	281
VNW(onbep,grad,stan,prenom,zonder,agr,basis)	990
VNW(onbep,grad,stan,prenom,zonder,agr,comp)	789
VNW(onbep,grad,stan,vrij,zonder,basis)	766
VNW(onbep,grad,stan,vrij,zonder,comp)	2102
VNW(onbep,grad,stan,vrij,zonder,sup)	287
VNW(onbep,pron,gen,vol,3p,ev)	19
VNW(onbep,pron,stan,vol,3o,ev)	1194
VNW(onbep,pron,stan,vol,3p,ev)	602
VNW(pers,pron,nomin,nadr,1,ev)	4
VNW(pers,pron,nomin,nadr,3m,ev,masc)	4
VNW(pers,pron,nomin,nadr,3v,ev,fem)	2
VNW(pers,pron,nomin,red,1,mv)	1224
VNW(pers,pron,nomin,red,2v,ev)	835
VNW(pers,pron,nomin,red,3,ev,masc)	87
VNW(pers,pron,nomin,red,3p,ev,masc)	510
VNW(pers,pron,nomin,vol,1,ev)	1319
VNW(pers,pron,nomin,vol,1,mv)	276
VNW(pers,pron,nomin,vol,2,getal)	5
VNW(pers,pron,nomin,vol,2b,getal)	2039
VNW(pers,pron,nomin,vol,2v,ev)	28
VNW(pers,pron,nomin,vol,3,ev,masc)	3759
VNW(pers,pron,nomin,vol,3p,mv)	584
VNW(pers,pron,nomin,vol,3v,ev,fem)	278
VNW(pers,pron,obl,nadr,3m,ev,masc)	10
VNW(pers,pron,obl,nadr,3p,mv)	3
VNW(pers,pron,obl,nadr,3v,getal,fem)	2
VNW(pers,pron,obl,red,3,ev,masc)	17

continued on next page

TAG	FREQUENCY
VNW(pers,pron,obl,vol,2v,ev)	14
VNW(pers,pron,obl,vol,3,ev,masc)	654
VNW(pers,pron,obl,vol,3,getal,fem)	115
VNW(pers,pron,obl,vol,3p,mv)	332
VNW(pers,pron,stan,nadr,2v,mv)	17
VNW(pers,pron,stan,red,3,ev,fem)	867
VNW(pers,pron,stan,red,3,ev,onz)	4452
VNW(pers,pron,stan,red,3,mv)	2056
VNW(pr,pron,obl,nadr,1,ev)	20
VNW(pr,pron,obl,nadr,1,mv)	2
VNW(pr,pron,obl,nadr,2,getal)	14
VNW(pr,pron,obl,nadr,2v,getal)	17
VNW(pr,pron,obl,red,1,ev)	123
VNW(pr,pron,obl,red,2v,getal)	118
VNW(pr,pron,obl,vol,1,ev)	107
VNW(pr,pron,obl,vol,1,mv)	246
VNW(pr,pron,obl,vol,2,getal)	14
VNW(recip,pron,gen,vol,persoon,mv)	12
VNW(recip,pron,obl,vol,persoon,mv)	421
VNW(refl,pron,obl,nadr,3,getal)	141
VNW(refl,pron,obl,red,3,getal)	2331
VNW(vb,adv-pron,obl,vol,3o,getal)	935
VNW(vb,det,stan,nom,met-e,zonder-n)	56
VNW(vb,det,stan,prenom,met-e,rest)	236
VNW(vb,det,stan,prenom,zonder,evon)	47
VNW(vb,pron,gen,vol,3m,ev)	18
VNW(vb,pron,gen,vol,3p,mv)	6
VNW(vb,pron,gen,vol,3v,ev)	1
VNW(vb,pron,stan,vol,3o,ev)	1142
VNW(vb,pron,stan,vol,3p,getal)	387
VZ(fin)	8096
VZ(init)	142147
VZ(versm)	880
WW(inf,nom,zonder,zonder-n)	2606
WW(inf,prenom,met-e)	2
WW(inf,vrij,zonder)	25665
WW(od,nom,met-e,mv-n)	32
WW(od,nom,met-e,zonder-n)	39
WW(od,prenom,met-e)	2224
WW(od,prenom,zonder)	378

continued on next page

TAG	FREQUENCY
WW(od,vrij,zonder)	602
WW(pv,conj,ev)	47
WW(pv,tgw,ev)	19853
WW(pv,tgw,met-t)	22286
WW(pv,tgw,mv)	16887
WW(pv,verl,ev)	20584
WW(pv,verl,mv)	7176
WW(vd,nom,met-e,mv-n)	262
WW(vd,nom,met-e,zonder-n)	162
WW(vd,prenom,met-e)	2743
WW(vd,prenom,zonder)	1553
WW(vd,vrij,zonder)	21846

Table B.9: CGN/D-Coi POS tags (`postag`) in LASSY Small (lexical nodes)

B.2.3 Data format

The annotations are encoded in the XML structure of the treebank. The data format of the LASSY treebank is Alpino-XML.[4]

The XML source code of the tree in Figure B.4 is given in Figure B.5. Every node in the tree is represented by a `node` element in the XML structure. The annotations presented in the tables above is included as XML attribute-value pairs in those nodes. The graphical representation only shows some of the features encoded in the XML data.

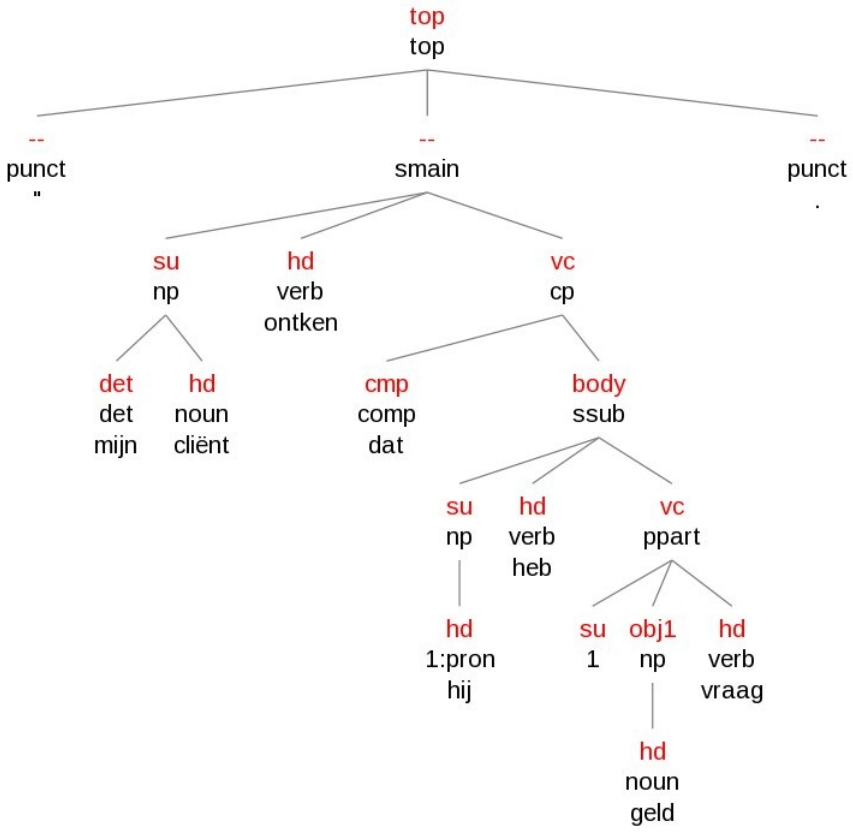


Figure B.4: Tree representation of a LASSY sentence (dpc-ind-001652-nl-sen.p.37.s.1)

[4] `http://www.let.rug.nl/vannoord/Lassy/alpino_ds.dtd`

```
<alpino_ds id="dpc-ind-001652-nl-sen.p.37.s.1" version="1.3">
  <node begin="0" cat="top" end="10" id="0" rel="top">
    <node begin="0" end="1" id="1" lemma="&quot;" pos="punct" postag="LET()" pt="let"
    rel="--" root="&quot;" word="&quot;"/>
    <node begin="1" cat="smain" end="9" id="2" rel="--">
      <node begin="1" cat="np" end="3" id="3" rel="su">
        <node begin="1" buiging="zonder" end="2" getal="ev" id="4" lemma="mijn"
        naamval="stan" npagr="agr" pdtype="det" persoon="1" pos="det" positie="prenom"
        postag="VNW(bez,det,stan,vol,1,ev,prenom,zonder,agr)" pt="vnw" rel="det"
        root="mijn" status="vol" vwtype="bez" word="Mijn"/>
        <node begin="2" end="3" genus="zijd" getal="ev" graad="basis" id="5"
        lemma="cliënt" naamval="stan" ntype="soort" pos="noun"
        postag="N(soort,ev,basis,zijd,stan)" pt="n" rel="hd" root="cliënt"
        word="cliënt"/>
      </node>
      <node begin="3" end="4" id="6" lemma="ontkennen" pos="verb"
      postag="WW(pv,tgw,met-t)" pt="ww" pvagr="met-t" pvtijd="tgw" rel="hd"
      root="ontken" word="ontkent" wvorm="pv"/>
      <node begin="4" cat="cp" end="9" id="7" rel="vc">
        <node begin="4" conjtype="onder" end="5" id="8" lemma="dat" pos="comp"
        postag="VG(onder)" pt="vg" rel="cmp" root="dat" word="dat"/>
        <node begin="5" cat="ssub" end="9" id="9" rel="body">
          <node begin="5" end="6" genus="masc" getal="ev" id="10" index="1" lemma="hij"
          naamval="nomin" pdtype="pron" persoon="3" pos="pron"
          postag="VNW(pers,pron,nomin,vol,3,ev,masc)" pt="vnw" rel="su" root="hij"
          status="vol" vwtype="pers" word="hij"/>
          <node begin="7" end="8" id="11" lemma="hebben" pos="verb"
          postag="WW(pv,tgw,met-t)" pt="ww" pvagr="met-t" pvtijd="tgw" rel="hd"
          root="heb" word="heeft" wvorm="pv"/>
          <node begin="5" cat="ppart" end="9" id="12" rel="vc">
            <node begin="5" end="6" id="13" index="1" rel="su"/>
            <node begin="6" end="7" genus="onz" getal="ev" graad="basis" id="14"
            lemma="geld" naamval="stan" ntype="soort" pos="noun"
            postag="N(soort,ev,basis,onz,stan)" pt="n" rel="obj1" root="geld"
            word="geld"/>
            <node begin="8" buiging="zonder" end="9" id="15" lemma="vragen" pos="verb"
            positie="vrij" postag="WW(vd,vrij,zonder)" pt="ww" rel="hd" root="vraag"
            word="gevraagd" wvorm="vd"/>
          </node>
        </node>
      </node>
    </node>
    <node begin="9" end="10" id="16" lemma="." pos="punct" postag="LET()" pt="let"
    rel="--" root="." word="."/>
  </node>
  <sentence>" Mijn cliënt ontkent dat hij geld heeft gevraagd .</sentence>
  <comments>
    <comment>Q#dpc-ind-001652-nl-sen.p.37.s.1|" Mijn cliënt ontkent dat hij geld heeft
    gevraagd .|</comment>
  </comments>
</alpino_ds>
```

Figure B.5: XML structure of a LASSY sentence (dpc-ind-001652-nl-sen.p.37.s.1)

Table B.10 contains all features that are included in the XML structures of the LASSY Small treebank. The first column lists all attributes that a node can contain, the second column provides a short description of the attributes, and the third column lists the possible values that each attribute can take.

The table is divided in two parts. The top part of the table describes the attributes that are assigned by the Alpino parser, except for the `lemma` and the (CGN/D-Coi) `postag` attributes, which are added to LASSY after parsing the treebank. Since the list of all CGN/D-Coi tags included in the treebank was presented in Table B.9, only three examples are included in this table. The complete list of CGN/D-Coi tags is provided in Van Eynde (2004, 2005).

The bottom part of the table contains attributes that refer to the information included in the CGN/D-Coi tags. After the CGN/D-Coi `postag` attributes were added to LASSY, they were split up. For example, a node that contains the attribute-value pair `postag="ADJ(nom,basis,met-e,mv-n)"`, contains the following attribute-value pairs as well: `pt="adj"`, `buiging="met-e"`, `graad="basis"`, `positie="nom"`, and `getal-n="mv-n"`. A detailed overview of those features can be found as well in Van Eynde (2004, 2005).

ATTRIBUTE	DESCRIPTION	POSSIBLE VALUES
rel	dependency relation	hdf/hd/cmp/sup/su/obj1/pobj1/obj2/se/pc/ vc/svp/predc /ld/me/predm/obcomp/mod/ body/det/app/whd/rhd/cnj/crd/ nucl/sat/ tag/dp/top/mwp/dlink/–
cat	syntactic category (non-leaf nodes)	smain/np/ppart/pp/ssub/inf/cp/du/ap/ advp/ti/rel/whrel/ whsub/conj/whq/ oti/ahi/detp/sv1/svan/mwu/top
index	index to co-index nodes	integer
begin	begin position	integer (the `begin` value of the first node is 0)
end	end position	integer (the `end` value of the first node is 1)
id	(unique) node identifier	integer
pos	(short) POS tag (leaf nodes)	adj/adv/comp/comparative/det/fixed/name/ noun/num/part/pp/prefix/prep/pron/punct/ tag/verb/vg
root	root form (leaf nodes)	text
word	word form, token (leaf nodes)	text
lemma	lemma	text
postag	CGN/D-Coi POS tag	e.g. ADJ(nom,basis,met-e,mv-n), BW(), WW(vd,vrij,zonder)
buiging	flexion	zonder/met-e/met-s
conjtype	conjunction type	neven/onder
dial	dialect word	dial
genus	gender	genus/zijd/masc/fem/onz
getal	number	getal/ev/mv
getal-n	number-n	zonder-n/mv-n
graad	degree	basis/comp/sup/dim
lwtype	article type	bep/onbep
naamval	case	stan/nomin/obl/bijz/gen/
npagr	NP agreement	agr/evon/rest/evz/mv/agr3/evmo/rest3/evf
ntype	noun type	soort/eigen
numtype	count type	hoofd/rang
pdtype	pronoun/determiner type	pron/adv-pron/det/grad
persoon	person	persoon/1/2/2v/2b/3/3p/3m/3v/3o
positie	position	prenom/nom/postnom/vrij
pt	short CGN/D-Coi POS tag	let/spec/bw/vg/lid/vnw/tw/ww/adj/n/tsw/vz
pvagr	finite verb agreement	ev/mv/met-t
pvtijd	finite verb tense	tgw/verl/conj
spectype	spec type	afgebr/onverst/vreemd/deeleigen/ meta/comment/achter/afk/symb/enof
status	pronoun/determiner reduction status	vol/red/nadr
vwtype	pronoun type	pr/pers/refl/recip/bez/vb/vrag/betr/ excl/aanw/onbep
vztype	preposition type	init/versm/fin
wvorm	verb vorm	pv/inf/od/vd

Table B.10: Features included in LASSY Small

Appendix C

XQuery script

The following XQuery script was used to find cluster creepers in two-verb clusters consisting of a finite verb and an infinitive. Comments are put between (: and :).

```
(: XPath extracts  V-final finite-infinitive clusters :)
(: in the LASSY small treebank :)

for $xp in db:open("LASSY_ID")/treebank/alpino_ds
//node[@cat="ssub" and node[@rel="hd" and @pt="ww" and @wvorm="pv"] and
node[@rel="vc" and @cat="inf" and node[@rel="hd" and @pt="ww"] and
not(node[@rel="vc" and (@cat="inf" or @cat="ti" or @cat="ppart" or @pt="ww")])]]

(: get sentence ID:)
let $sentenceid := ($xp/ancestor::alpino_ds/@id)

(: get sentence:)
let $sentence := ($xp/ancestor::alpino_ds/sentence)

(: get finite verb and infinitive :)
let $finite := ($xp/ node[@rel="hd" and @pt="ww" and @wvorm="pv"]/@word)
let $infinitive := ($xp/node[@rel="vc" and @cat="inf"]/
                    node[@rel="hd" and @pt="ww"]/@word)

(: get position of the finite verb and the infinitive :)
let $finiteposition := ($xp/ node[@rel="hd" and @pt="ww" and @wvorm="pv"]/@begin)
let $infinitiveposition := ($xp/node[@rel="vc" and @cat="inf"]/
                            node[@rel="hd" and @pt="ww"]/@begin)

(: get cluster creepers :)
(: finite - infinitive :)
let $creepers1 := ($xp/descendant::node[(number(@begin) > number($finiteposition))
and (number(@begin) < number($infinitiveposition))])
(: infinitive - finite :)
let $creepers2 := ($xp/descendant::node[(number(@begin) < number($finiteposition))
and (number(@begin) > number($infinitiveposition))])
```

```
(: only return constructions with cluster creepers :)
where ($creepers1 or $creepers2)

(: return sentences, verb cluster, and cluster creepers :)
return
if (number($finiteposition) < number($infinitiveposition))
then <match>{data($sentenceid)}#{data($sentence)}
#FINITE-INFINITIVE#{data($finite)}-{data($infinitive)}
#{data($creepers1/@word)}#{data($creepers1/@rel)}#{data($creepers1/@pt)}</match>

else
<match>{data($sentenceid)}#{data($sentence)}
#INFINITIVE-FINITE#{data($infinitive)}-{data($finite)}
#{data($creepers2/@word)}#{data($creepers2/@rel)}#{data($creepers2/@pt)}</match>
```

Bibliography

Abeillé, Anne (ed.). 2003. *Treebanks: Building and Using Parsed Corpora*. Dordrecht/Boston, MA/London: Kluwer Academic Publishers.

Abeillé, Anne, Danièle Godard, Philip Miller & Ivan A. Sag. 1998. French Bounded Dependencies. In Sergio Balari & Luca Dini (eds.), *Romance in HPSG*, 1–54. Stanford, CA: CSLI Publications.

Ackema, Peter. 2001. Colliding Complemetizers in Dutch: Another Syntactic OCP Effect. *Linguistic Inquiry* 32(4). 717–727.

Aelbrecht, Lobke. 2009. *You have the right to remain silent: The syntactic licensing of ellipsis*: Katholieke Universiteit Brussel dissertation.

Arfs, Mona. 2007. *Rood of groen? De interne woordvolgorde in tweeledige werkwoordelijke eindgroepen met een voltooid deelwoord en een hulpwerkwoord in bijzinnen*: Göteborg University dissertation.

Augustinus, Liesbeth & Frank Van Eynde. 2012. A Treebank-based Investigation of IPP-triggering Verbs in Dutch. In *Proceedings of the Eleventh International Workshop on Treebanks and Linguistic Theories (TLT11)*, 7–12. Lisbon: Edições Colibri.

Augustinus, Liesbeth & Frank Van Eynde. 2014. Looking for Cluster Creepers in Dutch Treebanks. Dat we ons daar nog kunnen mee bezig houden. *Computational Linguistics in the Netherlands Journal* 4. 149–170.

Augustinus, Liesbeth & Frank Van Eynde. in press. A Usage-based Typology of Dutch and German IPP verbs. *Leuvense Bijdragen* .

Augustinus, Liesbeth, Vincent Vandeghinste, Ineke Schuurman & Frank Van Eynde. 2013. Example-Based Treebank Querying with GrETEL - now also for Spoken Dutch. In *Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA)*, 423–428. Oslo: NEALT Proceedings Series 16.

Augustinus, Liesbeth, Vincent Vandeghinste & Frank Van Eynde. 2012. Example-Based Treebank Querying. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC)*, 3161–3167. Istanbul: European Language Resources Association (ELRA).

Bader, Markus & Tanja Schmid. 2009. Verb clusters in Colloquial German. *The Journal of Comparative Germanic Linguistics* 12(3). 175–228.

Barbiers, Sjef. 2005. Word order variation in three-verb clusters and the division of labour between generative linguistics and sociolinguistics. In Leonie Cornips & Karen Corrigan (eds.), *Syntax and Variation. Reconciling the Biological and the Social*, chap. 10, 233–264. Amsterdam/Philadelphia, PA: John Benjamins Publishing Company.

Barbiers, Sjef. 2008. Werkwoordclusters en de grammatica van de rechterperiferie. *Nederlandse Taalkunde* 13(2). 160–187.

Bech, Gunnar. 1955. *Studien über das deutsche Verbum infinitum* Det Kongelige Danske videnskabernes selskab. Historisk-filologiske Meddelelser 35:2. Copenhagen: Munksgaard.

Bennis, Hans. 1986. *Gaps and Dummies*: Katholieke Hogeschool Tilburg dissertation.

Bloem, Jelke, Arjen Versloot & Fred Weerman. 2014. Applying automatically parsed corpora to the study of language variation. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING)*, 1974–1984. Dublin.

Bouma, Gosse. 2000. Argument Realization and Dutch R-pronouns: Solving Bech's problem without Movement or Deletion. In Ronnie Cann, Claire Grover & Philip Miller (eds.), *Grammatical Interfaces in HPSG*, 51–76. Stanford, CA: CSLI Publications.

Bouma, Gosse. 2004. Treebank Evidence for the Analysis of PP-Fronting. In Sandra Kübler, Joakim Nivre, Erich Hinrichs & Holger Wunsch (eds.), *Proceedings of the Third Workshop on Treebanks and Linguistic Theories (TLT3)*, 15–26. Tübingen.

Bouma, Gosse & Geert Kloosterman. 2002. Querying Dependency Treebanks in XML. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC)*, Gran Canaria.

Bouma, Gosse & Geert Kloosterman. 2007. Mining Syntactically Annotated Corpora with XQuery. In *Proceedings of the Linguistic Annotation Workshop (ACL 07)*, 17–24. Prague.

Bouma, Gosse & Jennifer Spenader. 2009. The Distribution of Weak and Strong Object Reflexives in Dutch. In Frank Van Eynde, Anette Frank, Koenraad De Smedt & Gertjan van Noord (eds.), *Proceedings of the Seventh Workshop on Treebanks and Linguistic Theories (TLT7)* (LOT Occasional Series 12), 103–114. Utrecht.

Bouma, Gosse & Gertjan van Noord. 1998. Word Order Constraints on Verb Clusters in German and Dutch. In Erhard Hinrichs, Andreas Kathol & Tsuneko Nakazawa (eds.), *Complex Predicates in Nonderivational Syntax*, vol. 30 Syntax and Semantics, 43–72. New York: Academic Press.

Broekhuis, Hans. 2013. *Syntax of Dutch. Adpositions and Adpositional Phrases*. Amsterdam University Press.

Broekhuis, Hans & Norbert Corver. 2015. *Syntax of Dutch. Verbs and Verb Phrases*, vol. 2. Amsterdam University Press.

Chomsky, Noam. 1970. Remarks on Nominalization. In Roderick A. Jacobs & Peter S. Rosenbaum (eds.), *Readings in English Transformational Grammar*, 184–221. Waltham, MA: Ginn and Co.

Cinque, Guglielmo. 1999. *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. Oxford: Oxford University Press.

Coussé, Evie. 2006. De historische wortels van de volgordevariatie in het *hebben*-perfectum. *Taal en Tongval* 58. 250–276.

Coussé, Evie. 2008. *Motivaties voor volgordevariatie. Een diachrone studie van werkwoordsvolgorde in het Nederlands*: Ghent University dissertation.

De Sutter, Gert. 2005. *Rood, groen, corpus! Een taalgebruiksgebaseerde analyse van woordvolgordevariatie in tweeledige werkwoordelijke eindgroepen*: University of Leuven dissertation.

De Sutter, Gert. 2009. Towards a multivariate model of grammar: the case of word order variation in Dutch clause final verb clusters. In Andreas Dufter, Jürg Fleischer

& Guido Seiler (eds.), *Describing and Modeling Variation in Grammar*, 225–254. Berlin: Mouton de Gruyter.

den Besten, Hans & Jerold A. Edmondson. 1983. The Verbal Complex in Continental West Germanic. In Werner Abraham (ed.), *On the Formal Syntax of the Westgermania: Papers from the 3rd Groningen Grammar Talks (January 1981)*, 155–216. Amsterdam: Benjamins.

den Besten, Hans & Jean Rutten. 1989. On Verb Raising, Extraposition and Free Word Order in Dutch. In Dany Jaspers, Wim Klooster, Yves Putseys & Pieter Seuren (eds.), *Sentential Complementation and the Lexicon: Studies in Honour of Wim de Geest*, 41–56. Dordrecht: Foris.

den Besten, Hans, Jean Rutten, Tonjes Veenstra & Joop Veld. 1988. Verb Raising, Extrapositie en de Derde Constructie. Unpublished manuscript.

Draye, Luk & Joop van der Horst. 2006. Hebben/haben vs. zijn/sein in drieledige werkwoordsgroepen in het Duits en het Nederlands. In Matthias Hüning, Ulrike Vogl, Ton van der Wouden & Arie Verhagen (eds.), *Nederlands tussen Duits en Engels*, 229–242. Leiden: Stichting Neerlandistiek.

Duden. 2006. *DUDEN. Die Grammatik. Unentbehrlich für richtiges Deutsch.* Mannheim/Leipzig/Vienna/Zürich: Dudenverlag.

Evers, Arnold. 1975. *The Transformational Cycle in Dutch and German*: University of Utrecht dissertation.

Evers, Arnold. 2003. Verbal Clusters and Cluster Creepers. In Pieter A.M. Seuren & Gerard Kempen (eds.), *Verb Constructions in German and Dutch*, 43–89. Amsterdam/Philadelphia, PA: John Benjamins.

Evers, Arnold. 2004. Move Head. In Annemarie Kerkhoff, Joke de Lange & Oren Sadeh-Leicht (eds.), *Yearbook 2004*, 93–107. Utrecht Institute of Linguistics.

Fersøe, Hanne, Sussi Olsen, Costanza Navarretta & Bart Jongejan. 2006. Validation Report Corpus Gesproken Nederlands 1.0 Linguistic Validation. Tech. rep. Center for Sprogteknologi, University of Copenhagen.

Fleischer, Jürg. 2002. Preposition Stranding in German Dialects. In Sjef Barbiers, Leonie Cornips & Susanne van der Kleij (eds.), *Proceedings of the Workshop on Syntactic Microvariation, August 2000*, 116–151. Amsterdam: Meertens Instituut.

Ginzburg, Jonathan & Ivan Sag. 2000. *Interrogative investigations*. Stanford, CA: CSLI.

Haegeman, Liliane. 1988. Verb Projection Raising and the Multidimensional Analysis: Some Empirical Problems. *Linguistic Inquiry* 19(4). 671–683.

Haegeman, Liliane & Henk van Riemsdijk. 1986. Verb Projection Raising. Scope and the Typology of Rules Affecting Verbs. *Linguistic Inquiry* 17. 417–466.

Haeseryn, Walter, Kirsten Romijn, Geerts Geerts, Jaap de Rooij & Maarten van den Toorn. 1997. *Algemene Nederlandse Spraakkunst*. Groningen/Deurne: Martinus Nijhoff/Wolters Plantyn 2nd edn.

Hinrichs, Erhard & Tsuneko Nakazawa. 1989. Flipped Out: AUX in German. In *Papers from the 25th Regional Meeting of the Chicago Linguistic Society*, 193–202. Chicago, IL.

Hinrichs, Erhard & Tsuneko Nakazawa. 1990. Subcategorization and VP Structure in German. Department of Linguistics, University of Illinois.

Hinrichs, Erhard & Tsuneko Nakazawa. 1994. Linearizing AUXs in German Verbal Complexes. In John Nerbonne, Klaus Netter & Carl Pollard (eds.), *German in Head-Driven Phrase Structure Grammar*, 11–37. Stanford, CA: CSLI Publications.

Hinterhölzl, Roland. 1999. *Restructuring Infinitives and the Theory of Complementation*. Los Angeles: University of Southern California dissertation.

Hoeksema, Jack. 1980. Verbale verstrengeling ontstrengeld. *Spektator* 10. 221–249.

Hoekstra, Eric. 2010. On the interruption of Verb-Raising clusters by nonverbal material. In *Structure Preserved. Studies in Syntax for Jan Koster*, 175–184. Amsterdam: John Benjamins.

Hoekstra, Heleen, Michael Moortgat, Bram Renmans, Machteld Schouppe, Ineke Schuurman & Ton van der Wouden. 2003. *CGN Syntactische Annotatie*. 77p.

Höhle, Tilman. 1983. Topologische Felder. Ms, 78p.

IJbema, Aniek. 2002. *Grammaticalization and Infinitival Complements in Dutch*: Leiden University dissertation.

Jackendoff, Ray. 1977. *X-Bar Syntax: a Study of Phrase Structure*. Cambridge, MA: MIT Press.

Johnson, Mark. 1986. A GPSG Account of VP Structure in German. *Linguistics* 24(5). 871–882.

Jongejan, Bart, Sussi Olsen & Hanne Fersøe. 2011. Validation Report Lassy Corpora Linguistic Validation. Tech. rep. Center for Sprogteknologi, University of Copenhagen.

Jurafsky, Daniel & James H. Martin. 2009. *Speech and Language Processing*. New Jersey, USA: Pearson Education 2nd edn.

Kathol, Andreas. 2000. *Linear Syntax*. Oxford University Press.

Keenan, Edward L. & Bernard Comrie. 1977. Noun Phrase Accessibility and Universal Grammar. *Linguistic Inquiry* 8(1). 63–99.

Kim, Jong-Bok. 2009. Three Types of English Pseudo-passives: A Lexicalist Perspective. *Ene* 34(1). 1–24.

Kiss, Tibor. 1994. Obligatory Coherence: The Structure of German Modal Verb Constructions. In John Nerbonne, Klaus Netter & Carl Pollard (eds.), *German in Head-Driven Phrase Structure Grammar*, 71–107. Stanford, CA: CSLI Publications.

Kiss, Tibor. 1995. *Infinitive Komplementation. Neue Studien zum deutschen Verbum infinitum*. Tübingen: Niemeyer.

Klooster, Wim. 2001. *Grammatica van het hedendaags Nederlands. Een volledig overzicht*. Den Haag: Sdu Uitgevers.

Kloosterman, Geert. 2007. An Overview of the Alpino Treebank Tools. URL:http://www.let.rug.nl/vannoord/alp/Alpino/TreebankTools.html.

Kordoni, Valia & Gertjan van Noord. 2009. Passives in the Germanic Languages: the case of Dutch and German. *Groninger Arbeiten zur Germanistischen Linguistik (GAGL)* 49. 77–96.

Kübler, Sandra & Heike Zinsmeister. 2014. *Corpus Linguistics and Linguistically Annotated Corpora*. London/New York: Bloomsbury Academic.

Lezius, Wolfgang. 2002. TIGERSearch - ein Suchwerkzeug für Baumbanken. In Stephan Busemann (ed.), *Proceedings of KONVENS-02*, Saarbrücken.

Lüdeling, Anke & Merja Kytö (eds.). 2008. *Corpus Linguistics. An International Handbook*, vol. 1 & 2. Berlin: Mouton de Gruyter.

Manning, Christopher D. & Ivan A. Sag. 1999. Dissociations Between Argument Structure And Grammatical Relations. In Gert Webelhuth, Jean-Pierre Koenig & Andreas Kathol (eds.), *Lexical and Constructional Aspects of Linguistic Explanation* Studies in Constraint-Based Lexicalism, chap. 5, 63–78. Stanford, CA: CSLI Publications.

Miller, Philip H. & Ivan A. Sag. 1997. French Clitic Movement without Clitics or Movement. *Natural Language and Linguistic Theory* 15. 573–639.

Monachesi, Paola. 1998. Decomposing Italian Clitics. In Sergio Balari & Luca Dini (eds.), *Romance in HPSG*, 305–357. Stanford, CA: CSLI Publications.

Müller, Stefan. 1995. Scrambling in German – Extraction into the *Mittelfeld*. In Benjamin K. T'sou & Tom B. Y. Lai (eds.), *Proceedings of the 10th Pacific Asia Conference on Language, Information and Computation (PACLIC)*, 79–83. City University of Hong Kong.

Müller, Stefan. 2002. *Complex Predicates. Verbal Complexes, Resultative Constructions, and Particle Verbs in German*. Stanford, CA: CSLI.

Oostdijk, Nelleke, Wim Goedertier, Frank Van Eynde, Lou Boves, Jean-Pierre Martens, Michael Moortgat & Harald Baayen. 2002. Experiences from the Spoken Dutch Corpus Project. In Manuel Gonzalez Rodriguez & Carmen Paz Saurez Araujo (eds.), *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC2002)*, 340–347. Las Palmas.

Oostdijk, Nelleke, Martin Reynaert, Véronique Hoste & Ineke Schuurman. 2013. The Construction of a 500-Million-Word Reference Corpus of Contemporary Written Dutch. In Peter Spyns & Jan Odijk (eds.), *Essential Speech and Language Technology for Dutch. Results by the STEVIN programme*, 219–247. Springer.

Paulussen, Hans, Lieve Macken, Julia Truskina, Piet Desmet & Willy Vandeweghe. 2006. Dutch Parallel Corpus: a multifunctional and multilingual corpus. *Cahiers de l'Institut de Linguistique de Louvain, CILL* 32(1-4). 269–285.

Pauwels, Anita. 1953. *De plaats van hulpwerkwoord, verleden deelwoord en infinitief in de Nederlandse bijzin* (Werken uitgegeven door de Koninklijke Commissie voor Toponymie en Dialectologie 7). Leuven: Symons.

Plank, Frans. 2000. *Der Schritt vom Wege*: Form Alternations of the Prefix *k*- of the Resultative Participle in Bavarian. In Andreas Bittner, Dagmar Bittner & Klaus-Michael Köpcke (eds.), *Angemessene Strukturen: Systemorganisation in Phonologie, Morphologie und Syntax*, 17–30. Hildesheim, Germany: Olms.

Pollard, Carl. 1994. Toward a Unified Account of Passive in German. In John Nerbonne, Klaus Netter & Carl Pollard (eds.), *German in Head-Driven Phrase Structure Grammar*, 273–296. Stanford, CA: CSLI Publications.

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. University of Chicago Press.

Reape, Mike. 1993. *A Formal Theory of Word Order: A Case Study in West Germanic*: University of Edinburgh dissertation.

Rentier, Gerrit. 1993. Dutch Clitic Pronouns and Preposition Stranding. In *Computational Linguistics in the Netherlands. Papers from the third CLIN meeting*, 109–120. Tilburg.

Rentier, Gerrit. 1994. Dutch Cross Serial Dependencies in HPSG. In *Proceedings of COLING 94*, 818–822. Kyoto.

Rutten, Jean. 1991. *Infinitival Complements and Auxiliaries*: University of Amsterdam dissertation.

Sag, Ivan A. 2012. Sign-Based Construction Grammar: An Informal Synopsis. In Hans C. Boas & Ivan A. Sag (eds.), *Sign-Based Construction Grammar*, chap. 3, 69–202. Stanford, CA: CSLI.

Sag, Ivan A., Thomas Wasow & Emily M. Bender. 2003. *Syntactic Theory. A Formal Introduction*. Stanford, CA: CSLI Publications 2nd edn.

Salzmann, Martin. 2013. New arguments for verb cluster formation at PF and a right-branching VP. Evidence from verb doubling and cluster penetrability. *Linguistic Variation* 13(1). 81–132.

Schmid, Tanja. 2005. *Infinitival Syntax. Infinitivus Pro Participio as a repair strategy.* Amsterdam/Philadelphia, PA: John Benjamins.

Smessaert, Hans, William Van Belle & Ingrid Van Canegem-Ardijns. 2014. Adpositional constructions of location and motion in Dutch. In Nicole Delbecque, Karen Lahousse & Willy Van Langendonck (eds.), *Non-Nuclear Cases*, vol. 6 Case and grammatical relations across languages, 127–171. Amsterdam/Philadelphia, PA: John Benjamins.

Steedman, M. 1985. Dependency and Coordination in the Grammar of Dutch and English. *Language* 61(3). 523–568.

Steedman, Mark. 2000. *The Syntactic Process*. Cambridge, MA: MIT Press.

Ter Beek, Janneke. 2008. *Restructuring and Infinitival Complements in Dutch*: University of Groningen dissertation.

Uszkoreit, Hans. 1987a. Linear Precedence in Discontinuous Constituents: Complex Fronting in German. In Geoffrey Huck & Almerindo Ojeda (eds.), *Discontinuous constituency*, vol. 20 Syntax and Semantics, 405–425. Orlando, FL/London: Academic Press.

Uszkoreit, Hans. 1987b. *Word Order and Constituent Structure in German* (Lecture Notes 8). Stanford, CA: CSLI Publications.

Van den Bosch, Antal, Bertjan Busser, Walter Daelemans & Sander Canisius. 2007. An efficient memory-based morphosyntactic tagger and parser for Dutch. In Peter Dirix, Ineke Schuurman, Vincent Vandeghinste & Frank Van Eynde (eds.), *Selected Papers from the 17th Computational Linguistics in the Netherlands (CLIN) Meeting*, 191–206. Utrecht: LOT.

van der Beek, Leonoor. 2005. *Topics in Corpus-Based Syntax*: University of Groningen dissertation.

van der Beek, Leonoor, Gosse Bouma, Rob Malouf & Gertjan van Noord. 2002. The Alpino Dependency Treebank. In *Computational Linguistics in the Netherlands 2001*, Amsterdam/Atlanta, GA: Rodopi.

Van Eynde, Frank. 1998. The Immediate Dominance Schemata of HPSG. A Deconstruction and a Reconstruction. In Peter-Arno Coppen, Hans van Halteren & Lisanne Teunissen (eds.), *Computational Linguistics in the Netherlands 1997. Selected Papers from the Eighth CLIN Meeting*, 119–133. Amsterdam/Atlanta, GA: Rodopi.

Van Eynde, Frank. 1999. Major and Minor Pronouns in Dutch. In Gosse Bouma, Erhard Hinrichs, Geert-Jan M. Kruijff & Richard Oehrle (eds.), *Constraints and Resources in Natural Language Syntax and Semantics*, 137–152. Stanford, CA: CSLI Publications.

Van Eynde, Frank. 2004. Part of Speech Tagging en Lemmatisering van het Corpus Gesproken Nederlands. 87 pages.

Van Eynde, Frank. 2005. Part of Speech Tagging en Lemmatisering van het D-Coi Corpus. 88 pages.

Van Eynde, Frank. 2006. NP-internal agreement and the structure of the noun phrase. *Journal of Linguistics* 42. 139–186.

Van Eynde, Frank. 2009. A Treebank-driven investigation of predicative complements in Dutch. In Barbara Plank, Erik Tjong Kim Sang & Tim Van de Cruys (eds.), *Computational Linguistics in the Netherlands 2009. Selected papers from the nineteenth CLIN meeting* LOT Occasional Series 14, 131–145. Utrecht.

Van Eynde, Frank. 2015. *Predicative Constructions. From the Fregean to a Montagovian Treatment*. Stanford, CA: CSLI Publications.

Van Eynde, Frank & Liesbeth Augustinus. 2013. Why and How to Differentiate Complement Raising from Subject Raising in Dutch. In Stefan Müller (ed.), *Proceedings of the 20th International Conference on Head-Driven Phrase Structure Grammar*, 222–242. Stanford, CA: CSLI Publications.

Van Eynde, Frank & Liesbeth Augustinus. 2014. Complement Raising, Extraction and Adpostion Stranding in Dutch. In Stefan Müller (ed.), *Proceedings of the 21th International Conference on Head-Driven Phrase Structure Grammar*, 156–175. Stanford, CA: CSLI Publications.

Van Eynde, Frank, Liesbeth Augustinus, Ineke Schuurman & Vincent Vandeghinste. in press. *Hebben* of *zijn* bij IPP's. *Leuvense Bijdragen* .

van Noord, Gertjan. 2006. At Last Parsing Is Now Operational. In Piet Mertens, Cédrick Fairon, Anne Dister & Patrick Watrin (eds.), *TALN 2006. Verbum Ex Machina. Actes de la 13ème conference sur le Traitement Automatique des Langues Naturelles*, 20–42.

van Noord, Gertjan & Gosse Bouma. 1997. Dutch Verb Clustering without Verb Clusters. In Patrick Blackburn & Maarten De Rijke (eds.), *Specifying Syntactic Structures*, 213–243. Stanford, CA: CSLI.

van Noord, Gertjan, Gosse Bouma, Frank Van Eynde, Daniël de Kok, Jelmer van der Linde, Ineke Schuurman, Erik Tjong Kim Sang & Vincent Vandeghinste. 2013. Large Scale Syntactic Annotation of Written Dutch: Lassy. In Peter Spyns & Jan Odijk (eds.), *Essential Speech and Language Technology for Dutch: Results by the STEVIN programme*, 147–164. Springer.

van Noord, Gertjan, Ineke Schuurman & Gosse Bouma. 2011. *Lassy Syntactische Annotatie, Revision 19455*. www.let.rug.nl/vannoord/Lassy/sa-man_lassy.pdf.

Van Riemsdijk, Henk C. 1978. *A Case Study in Syntactic Markedness. The Binding Nature of Prepositional Phrases*: University of Amsterdam dissertation.

Warner, Anthony R. 2000. English auxiliaries without lexical rules. In Robert D. Borsley (ed.), *The Nature and Function of Syntactic Categories*, vol. 32 Syntax and Semantics, 167–220. San Diego, CA/London: Academic Press.

Wurmbrand, Susi. 2001. Bibliography on verb clusters and verb (projection) raising. `http://wurmbrand.uconn.edu/Bibliographies/vc-bib.html`. Last updated February, 2006.

Wurmbrand, Susi. 2004. West Germanic verb clusters: The empirical domain. In Katalin É. Kiss & Henk van Riemsdijk (eds.), *Verb clusters: A study of Hungarian, German, and Dutch*, 43–85. Amsterdam/Philadelphia, PA: John Benjamins.

Wurmbrand, Susi. 2005. Verb Clusters, Verb Raising, and Restructuring. In Martin Everaert & Henk van Riemsdijk (eds.), *The Blackwell Companion to Syntax*, vol. V, chap. 75, 229–343. Oxford: Blackwell.

Wurmbrand, Susi. in press. Verb Clusters, Verb Raising, and Restructuring. In Martin Everaert & Henk van Riemsdijk (eds.), *The Blackwell Companion to Syntax*, Oxford: Blackwell 2nd edn.

Zwart, Jan-Wouter. 1994. Dutch is Head Initial. *The Linguistic Review* 11(3/4). 377–406.

Zwart, Jan-Wouter. 1996. Verb clusters in Continental West Germanic dialects. In James R. Black & Virginia Motapanyane (eds.), *Microparametric Syntax and Dialect Variation*, 229–258. Amsterdam: Benjamins.

Zwart, Jan-Wouter. 2007. Some notes on the origin and distribution of the IPP-effect. *Groninger Arbeiten zur Germanistischen Linguistik (GAGL)* 45. 77–99.

Zwart, Jan-Wouter. 2011. *The Syntax of Dutch*. Cambridge University Press.

Samenvatting

De vorming van werkwoordclusters is een alom bekend fenomeen binnen de Nederlandse syntaxis. In constructies als *Ik denk dat ik Cecilia het nijlpaard heb zien voeren* vormen de werkwoorden aan het einde van de zin een cluster, waardoor ze gescheiden worden van hun niet-werkwoordelijke argumenten. In principe is een werkwoordcluster ondoordringbaar: in het algemeen kunnen er geen niet-werkwoordelijke elementen tussen de werkwoorden staan. Het bovenstaande voorbeeld bevat bovendien nog een opmerkelijk fenomeen: het werkwoord *zien* wordt geselecteerd door het perfectief werkwoord *hebben*, maar neemt de vorm aan van een infinitief in plaats van een deelwoord. Dergelijke constructies staan bekend als *Infinitivus pro Participio* of IPP.

Werkwoordclusters en verwante fenomenen zoals IPP zijn uitgebreid bestudeerd in zowel de descriptieve en theoretische taalkunde als de corpuslinguïstiek. Toch zijn er nog heel wat aspecten met betrekking tot werkwoordclusters die verder onderzoek vergen. Het onderzoek dat wordt gepresenteerd in deze dissertatie tracht een aantal van deze vragen te beantwoorden. Daarbij wordt ook onderzocht hoe een corpusonderzoek tot nieuwe inzichten kan leiden voor de syntactische analyse van werkwoordclusters.

De dissertatie bestaat uit drie delen: een literatuurstudie, een corpusonderzoek, en een theoretische analyse.

Deel I bespreekt hoe werkwoordclusters beschreven en geanalyseerd worden binnen de **descriptieve en theoretische literatuur**. Hoofdstuk 1 geeft een definitie van werkwoordclusters op basis van de literatuur, en bespreekt verschillende fenomenen die vaak gerelateerd worden aan clustervorming. De belangrijkste zijn het IPP-effect, de doorbreking van de cluster door niet-werkwoordelijke elementen, en woordvolgordevariatie binnen de cluster.

De literatuurstudie behandelt de volgende onderzoeksvragen:

- Welke zijn de clusterende werkwoorden in het Nederlands?
- In welke gevallen is clustervorming verplicht en in welke gevallen is het optioneel?
- Wat is de relatie tussen werkwoordclustering en het IPP-effect?
- Welke soorten woordvolgordevariatie komen voor in werkwoordclusters?
- In welke mate komt doorbreking van de cluster voor? Onder welke voorwaarden kan het fenomeen zich voordoen?

Uit de literatuurstudie blijkt dat de verzameling van clusterende werkwoorden niet uniform gedefinieerd is. Verder blijkt dat constructies met een optioneel IPP-werkwoord vaak ambigu zijn tussen clusterende en niet-clusterende constructies.

Hoofdstuk 2 bespreekt hoe werkwoordclusters geanalyseerd worden in transformationele benaderingen. De eerste theoretische analyses van Nederlandse werkwoordclusters werden geformuleerd in dit framework. Daarnaast is de terminologie m.b.t. clustervorming in de descriptieve literatuur en in andere frameworks vaak gebaseerd op begrippen uit de transformationele literatuur. De transformationele analyses vertrekken vanuit de assumptie dat werkwoorden een syntactisch en semantisch geheel vormen met hun argumenten. Door transformaties toe te passen op die structuur, kunnen constructies met werkwoordclusters afgeleid worden.

Hoofdstuk 3 bespreekt de belangrijkste monostratale analyses van werkwoordclusters. De klemtoon ligt daarbij op de analyses geformuleerd in het framework van Head-driven Phrase Structure Grammar (HPSG), aangezien dat framework ook gehanteerd wordt voor de nieuwe analyse. In tegenstelling tot transformationele grammatica, worden werkwoordclusters binnen HPSG niet geanalyseerd door middel van verplaatsing, maar door de overerving van niet-gesatureerde valentielijsten. Een analyse die veel navolging heeft gekregen, maakt gebruik van het *argument inheritance* principe, voorgesteld door Hinrichs & Nakazawa (1994). Het breidt de analyse van subjectraising uit naar een algemener principe, waarin subjecten en complementen op een analoge manier verheven worden. De analyse is oorspronkelijk geformuleerd voor het Duits, maar is ook gebruikt om Nederlandse werkwoordsclusters te analyseren.

Deel II van deze dissertatie beschrijft het **corpusonderzoek**. Door gebruik te maken van syntactisch geannoteerde corpora of *treebanks* wordt onderzocht of de fenomenen die besproken werden in de literatuurstudie ook voorkomen in niet-geëliciteerde

data. Hoofdstuk 4 bespreekt de data en de methodologie die toegepast werd voor het corpusonderzoek. Hoofdstuk 5 beschrijft en bespreekt de resultaten die voortkomen uit dat onderzoek. De belangrijkste fenomenen die aan bod komen zijn de woordvolgordevariatie in werkwoordclusters, de identificatie van clusterende werkwoorden, het IPP-effect en de doorbreking van de cluster. Constructies met een *te*-infinitief krijgen speciale aandacht, aangezien ze vaak onderbelicht of zelfs geïgnoreerd worden in het onderzoek naar werkwoordclusters.

Het onderzoek naar woordvolgordevariatie in de cluster bevestigt de generalisaties m.b.t. het Standaardnederlands, geformuleerd in Broekhuis & Corver (2015): korte infinitieven en *te*-infinitieven volgen steeds op het selecterend werkwoord (bv. *heeft kunnen komen*, *heeft proberen te komen*), terwijl deelwoorden op andere plaatsen in de cluster kunnen voorkomen (bv. *moet hebben gelezen*, *moet gelezen hebben*, *gelezen moet hebben*). De treebankdata bevatten slechts enkele uitzonderingen, voornamelijk gevallen waarin een finiet modaal of aspectueel werkwoord een infinitief selecteert (bv. *lezen moet*).

In spreektaal en informeel taalgebruik (voornamelijk in Vlaanderen) komen er ook constructies voor waarin het perfectieve werkwoord volgt op een IPP werkwoord en het complement daarvan (bv. *kunnen komen heeft*). De gevallen waarin een *te*-infinitief voor het selecterende werkwoord staat, blijken geen voorbeelden van werkwoordsclusters te zijn. Dergelijke *te*-infinitieven gedragen zich eerder als predicatieve complementen, of vormen een onderdeel van een idiomatische uitdrukking.

Naast de woordvolgordevariatie werden de clusterende werkwoorden en de werkwoorden die het IPP-effect kunnen vertonen geïdentificeerd. Niet alle werkwoorden die in de literatuur besproken worden komen voor in de treebanks als gevolg van 'data sparseness'. Om een datagebaseerde lijst te bekomen, werd de typologie die uit de treebankdata voortkomt aangevuld met attestaties van het web.

Een laatste casestudy beschrijft het onderzoek naar de doorbreking van de cluster (*cluster creeping*). Zoals verwacht, komt het fenomeen vaker voor in gesproken dan in geschreven data. Het treebankonderzoek resulteert in een classificatie van elementen die de cluster kunnen doorbreken (*cluster creepers*), gebaseerd op hun lexicale categorie en syntactische functie.

Deel III beschrijft een **nieuwe analyse van Nederlandse werkwoordclusters**, geformuleerd in HPSG. In hoofdstuk 6 wordt aangetoond dat het *argument inheritance* mechanisme geen accurate analyse biedt voor Nederlandse werkwoordsclusters. Zo kan complementraising ook voorkomen zonder subjectraising in clusters met subject-

controle werkwoorden (bv. *dat hij dat boek al vaak heeft proberen te lezen*). Daarnaast is het mechanisme niet compatibel met de principes van de bindingstheorie, en met de lexicale regel voor de vorming van het passief.

De alternatieve analyse die wordt voorgesteld, behandelt raising van subjecten en complementen als aparte fenomenen. In het geval van subject raising worden de subjectvereisten van het geselecteerde werkwoord op de gebruikelijke manier gedeeld met het subject van het selecterende werkwoord. De complementvereisten worden niet gedeeld tussen het selecterende werkwoord en het geselecteerde werkwoord, maar worden gedeeld tussen het werkwoordelijk complement en de moederknoop.

De analyse is gebaseerd op intra- en interlinguïstische observaties, en wordt zoveel mogelijk empirisch onderbouwd op basis van voorbeelden uit de treebanks. In het Nederlands komen subject- en complementraising niet noodzakelijk samen voor, cf. het voorbeeld met *proberen*. Daarnaast zijn er talen die subjectraising vertonen, maar geen complementraising (bv. Engels).

Tenslotte wordt in hoofdstuk 7 aangetoond hoe de voorgestelde analyse niet louter werkwoordclusters kan behandelen, maar ook kan toegepast worden op andere fenomenen, zoals raising uit niet-verbale complementen, bv. adjectivale en prepositionele complementen. Raising uit prepositionele complementen leidt tot adpositiestranding. Een voorbeeld is de raising van *daar*, dat afgesplitst is van de adpositie *aan* in *dat hij daar nog niet aan gedacht had*. Adpositiestranding wordt verder uitgewerkt in het laatste deel van deze dissertatie, waarin wordt aangetoond dat complement raising mogelijk is uit P-finale PP's, maar niet uit P-initiële PP's. Verder wordt aangetoond dat complementraising moet onderscheiden worden van complementextractie.